



---

# Probability

---



Leo Breiman

---

C · L · A · S · S · I · C · S

---

In Applied Mathematics

**siam**

7

---



---



# Probability



---



SIAM's Classics in Applied Mathematics series consists of books that were previously allowed to go out of print. These books are republished by SIAM as a professional service because they continue to be important resources for mathematical scientists.

### **Editor-in-Chief**

Robert E. O'Malley, Jr., *University of Washington*

### **Editorial Board**

Richard A. Brualdi, *University of Wisconsin-Madison*

Herbert B. Keller, *California Institute of Technology*

Andrzej Z. Manitius, *George Mason University*

Ingram Olkin, *Stanford University*

Stanley Richardson, *University of Edinburgh*

Ferdinand Verhulst, *Mathematisch Instituut, University of Utrecht*

### **Classics in Applied Mathematics**

C. C. Lin and L. A. Segel, *Mathematics Applied to Deterministic Problems in the Natural Sciences*

Johan G. F. Belinfante and Bernard Kolman, *A Survey of Lie Groups and Lie Algebras with Applications and Computational Methods*

James M. Ortega, *Numerical Analysis: A Second Course*

Anthony V. Fiacco and Garth P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*

F. H. Clarke, *Optimization and Nonsmooth Analysis*

George F. Carrier and Carl E. Pearson, *Ordinary Differential Equations*

Leo Breiman, *Probability*

R. Bellman and G. M. Wing, *An Introduction to Invariant Imbedding*

Abraham Berman and Robert J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*

Olvi L. Mangasarian, *Nonlinear Programming*

\*Carl Friedrich Gauss, *Theory of the Combination of Observations Least Subject to Errors: Part One, Part Two, Supplement*. Translated by G. W. Stewart

Richard Bellman, *Introduction to Matrix Analysis*

U. M. Ascher, R. M. M. Mattheij, and R. D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*

K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*

Charles L. Lawson and Richard J. Hanson, *Solving Least Squares Problems*

J. E. Dennis, Jr. and Robert B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*

Richard E. Barlow and Frank Proschan, *Mathematical Theory of Reliability*

\*First time in print.

## Classics in Applied Mathematics (continued)

Cornelius Lanczos, *Linear Differential Operators*

Richard Bellman, *Introduction to Matrix Analysis, Second Edition*

Beresford N. Parlett, *The Symmetric Eigenvalue Problem*

Richard Haberman, *Mathematical Models: Mechanical Vibrations, Population Dynamics, and Traffic Flow*

Peter W. M. John, *Statistical Design and Analysis of Experiments*

Tamer Başar and Geert Jan Olsder, *Dynamic Noncooperative Game Theory, Second Edition*

Emanuel Parzen, *Stochastic Processes*

Petar Kokotović, Hassan K. Khalil, and John O'Reilly, *Singular Perturbation Methods in Control: Analysis and Design*

Jean Dickinson Gibbons, Ingram Olkin, and Milton Sobel, *Selecting and Ordering Populations: A New Statistical Methodology*

James A. Murdock, *Perturbations: Theory and Methods*

Ivar Ekeland and Roger Témam, *Convex Analysis and Variational Problems*

Ivar Stakgold, *Boundary Value Problems of Mathematical Physics, Volumes I and II*

J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*

David Kinderlehrer and Guido Stampacchia, *An Introduction to Variational Inequalities and Their Applications*

F. Natterer, *The Mathematics of Computerized Tomography*

Avinash C. Kak and Malcolm Slaney, *Principles of Computerized Tomographic Imaging*

R. Wong, *Asymptotic Approximations of Integrals*

O. Axelsson and V. A. Barker, *Finite Element Solution of Boundary Value Problems: Theory and Computation*

David R. Brillinger, *Time Series: Data Analysis and Theory*

Joel N. Franklin, *Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems*

Philip Hartman, *Ordinary Differential Equations, Second Edition*

Michael D. Intriligator, *Mathematical Optimization and Economic Theory*

Philippe G. Ciarlet, *The Finite Element Method for Elliptic Problems*

Jane K. Cullum and Ralph A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol. I: Theory*

M. Vidyasagar, *Nonlinear Systems Analysis, Second Edition*

Robert Mattheij and Jaap Molenaar, *Ordinary Differential Equations in Theory and Practice*

Shanti S. Gupta and S. Panchapakesan, *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*



*This page intentionally left blank*



---

# Probability

---



Leo Breiman  
University of California, Berkeley

**siam**

Society for Industrial and Applied Mathematics  
Philadelphia

Copyright ©1992 by the Society for Industrial and Applied Mathematics.

This SIAM edition is an unabridged, corrected republication of the work first published by Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1968.

10 9 8 7 6 5

All rights reserved. Printed in the United States of America. No part of this book may be reproduced, stored, or transmitted in any manner without the written permission of the publisher. For information, write to the Society for Industrial and Applied Mathematics, 3600 University City Science Center, Philadelphia, PA 19104-2688.

**Library of Congress Cataloging-in-Publication Data**

Breiman, Leo

Probability / Leo Breiman.

p. cm. -- (Classics in applied mathematics ; 7)

Originally published: Reading, Mass. : Addison-Wesley Pub. Co.,  
1968. (Addison-Wesley series in statistics)

Includes bibliographical references and index.

ISBN 0-89871-296-3

I. Probabilities. I. Title. II. Series.

QA273.B864 1992

519.2--dc20

92-1381

# Preface to the Classic Edition

This is the first of four books I have written; the one I worked the hardest on; and the one I am fondest of. It marked my goodbye to mathematics and probability theory. About the time the book was written, I left UCLA to go into the world of applied statistics and computing as a full-time freelance consultant.

The book went out of print well over ten years ago, but before it did a generation of statisticians, engineers, and mathematicians learned graduate probability theory from its pages. Since the book became unavailable, I have received many calls asking where it could be bought and then for permission to copy part or all of it for use in graduate probability courses.

These reminders that the book was not forgotten saddened me and I was delighted when SIAM offered to republish it in their Classics Series. The present edition is the same as the original except for the correction of a few misprints and errors, mainly minor.

After the book was out for a few years it became commonplace for a younger participant at some professional meeting to lean over toward me and confide that he or she had studied probability out of my book. Lately, this has become rarer and the confiders older. With republication, I hope that the age and frequency trends will reverse direction.

Leo Breiman  
University of California, Berkeley  
January, 1992

*This page intentionally left blank*

# Preface

A few years ago I started a book by first writing a very extensive preface. I never finished that book and resolved that in the future I would write first the book and then the preface. Having followed this resolution I note that the result is a desire to be as brief as possible.

This text developed from an introductory graduate course and seminar in probability theory at UCLA. A prerequisite is some knowledge of real variable theory, such as the ideas of measure, measurable functions, and so on. Roughly, the first seven chapters of *Measure Theory* by Paul Halmos [64] is sufficient background. There is an appendix which lists the essential definitions and theorems. This should be taken as a rapid review or outline for study rather than as an exposition. No prior knowledge of probability is assumed, but browsing through an elementary book such as the one by William Feller [59, Vol. I], with its diverse and vivid examples, gives an excellent feeling for the subject.

Probability theory has a right and a left hand. On the right is the rigorous foundational work using the tools of measure theory. The left hand “thinks probabilistically,” reduces problems to gambling situations, coin-tossing, motions of a physical particle. I am grateful to Michel Loève for teaching me the first side, and to David Blackwell, who gave me the flavor of the other.

David Freedman read through the entire manuscript. His suggestions resulted in many substantial revisions, and the book has been considerably improved by his efforts. Charles Stone worked hard to convince me of the importance of analytic methods in probability. The presence of Chapter 10 is largely due to his influence, and I am further in his debt for reading parts of the manuscript and for some illuminating conversations on diffusion theory.

Of course, in preparing my lectures, I borrowed heavily from the existing books in the field and the finished product reflects this. In particular, the books by M. Loève [108], J. L. Doob [39], E. B. Dynkin [43], and K. Ito and H. P. McKean [76] were significant contributors.

Two students, Carl Maltz and Frank Kontrovich, read parts of the manuscript and provided lists of mistakes and unreadable portions. Also, I was blessed by having two fine typists, Louise Gaines and Ruth Goldstein, who rose above mere patience when faced with my numerous revisions of the “final draft.” Finally, I am grateful to my many nonmathematician friends who continually asked when I was going to finish “that thing,” in voices that could not be interminably denied.

Leo Breiman  
Topanga, California  
January, 1968

*This page intentionally left blank*

# Contents

## Chapter 1 Introduction

- 1  $n$  independent tosses of a fair coin 1
- 2 The “law of averages” 1
- 3 The bell-shaped curve enters (fluctuation theory) 7
- 4 Strong form of the “law of averages” 11
- 5 An analytic model for coin-tossing 15
- 6 Conclusions 17

## Chapter 2 Mathematical Framework

- 1 Introduction 19
- 2 Random vectors 20
- 3 The distribution of processes 21
- 4 Extension in sequence space 23
- 5 Distribution functions 25
- 6 Random variables 29
- 7 Expectations of random variables 31
- 8 Convergence of random variables 33

## Chapter 3 Independence

- 1 Basic definitions and results 36
- 2 Tail events and the Kolmogorov zero-one law 40
- 3 The Borel-Cantelli lemma 41
- 4 The random signs problem 45
- 5 The law of pure types 49
- 6 The law of large numbers for independent random variables 51
- 7 Recurrence of sums 53
- 8 Stopping times and equidistribution of sums 58
- 9 Hewitt-Savage zero-one law 63

## Chapter 4 Conditional Probability and Conditional Expectation

- 1 Introduction 67
- 2 A more general conditional expectation 73
- 3 Regular conditional probabilities and distributions 77



## **Chapter 5      Martingales**

- 1    Gambling and gambling systems    82
- 2    Definitions of martingales and submartingales    83
- 3    The optional sampling theorem    84
- 4    The martingale convergence theorem    89
- 5    Further martingale theorems    91
- 6    Stopping times    95
- 7    Stopping rules    98
- 8    Back to gambling    101

## **Chapter 6      Stationary Processes and the Ergodic Theorem**

- 1    Introduction and definitions    104
- 2    Measure-preserving transformations    106
- 3    Invariant sets and ergodicity    108
- 4    Invariant random variables    112
- 5    The ergodic theorem    113
- 6    Converses and corollaries    116
- 7    Back to stationary processes    118
- 8    An application    120
- 9    Recurrence times    122
- 10   Stationary point processes    125

## **Chapter 7      Markov Chains**

- 1    Definitions    129
- 2    Asymptotic stationarity    133
- 3    Closed sets, indecomposability, ergodicity    135
- 4    The countable case    137
- 5    The renewal process of a state    138
- 6    Group properties of states    141
- 7    Stationary initial distributions    143
- 8    Some examples    145
- 9    The convergence theorem    150
- 10   The backward method    153

## **Chapter 8      Convergence in Distribution and the Tools Thereof**

- 1    Introduction    159
- 2    The compactness of distribution functions    160
- 3    Integrals and  $\mathcal{D}$ -convergence    163
- 4    Classes of functions that separate    165
- 5    Translation into random-variable terms    166
- 6    An application of the foregoing    167
- 7    Characteristic functions and the continuity theorem    170

8	The convergence of types theorem	174
9	Characteristic functions and independence	175
10	Fourier inversion formulas	177
11	More on characteristic functions	179
12	Method of moments	181
13	Other separating function classes	182

## **Chapter 9      The One-Dimensional Central Limit Problem**

1	Introduction	185
2	Why normal?	185
3	The nonidentically distributed case	186
4	The Poisson convergence	188
5	The infinitely divisible laws	190
6	The generalized limit problem	195
7	Uniqueness of representation and convergence	196
8	The stable laws	199
9	The form of the stable laws	200
10	The computation of the stable characteristic functions	204
11	The domain of attraction of a stable law	207
12	A coin-tossing example	213
13	The domain of attraction of the normal law	214

## **Chapter 10      The Renewal Theorem and Local Limit Theorem**

1	Introduction	216
2	The tools	216
3	The renewal theorem	218
4	A local central limit theorem	224
5	Applying a Tauberian theorem	227
6	Occupation times	229

## **Chapter 11      Multidimensional Central Limit Theorem and Gaussian Processes**

1	Introduction	233
2	Properties of $\mathcal{N}_k$	234
3	The multidimensional central limit theorem	237
4	The joint normal distribution	238
5	Stationary Gaussian process	241
6	Spectral representation of stationary Gaussian processes	242
7	Other problems	246

## **Chapter 12      Stochastic Processes and Brownian Motion**

- 1    Introduction    248
- 2    Brownian motion as the limit of random walks    251
- 3    Definitions and existence    251
- 4    Beyond the Kolmogorov extension    254
- 5    Extension by continuity    255
- 6    Continuity of Brownian motion    257
- 7    An alternative definition    259
- 8    Variation and differentiability    261
- 9    Law of the iterated logarithm    263
- 10   Behavior at  $t = \infty$     265
- 11   The zeros of  $X(t)$     267
- 12   The strong Markov property    268

## **Chapter 13      Invariance Theorems**

- 1    Introduction    272
- 2    The first-exit distribution    273
- 3    Representation of sums    276
- 4    Convergence of sample paths of sums to Brownian motion paths    278
- 5    An invariance principle    281
- 6    The Kolmogorov-Smirnov statistics    283
- 7    More on first-exit distributions    287
- 8    The law of the iterated logarithm    291
- 9    A more general invariance theorem    293

## **Chapter 14      Martingales and Processes with Stationary, Independent Increments**

- 1    Introduction    298
- 2    The extension to smooth versions    298
- 3    Continuous parameter martingales    300
- 4    Processes with stationary, independent increments    303
- 5    Path properties    306
- 6    The Poisson process    308
- 7    Jump processes    310
- 8    Limits of jump processes    312
- 9    Examples    316
- 10   A remark on a general decomposition    318

## **Chapter 15     Markov Processes, Introduction and Pure Jump Case**

- 1    Introduction and definitions    319
- 2    Regular transition probabilities    320
- 3    Stationary transition probabilities    322
- 4    Infinitesimal conditions    324
- 5    Pure jump processes    328
- 6    Construction of jump processes    332
- 7    Explosions    336
- 8    Nonuniqueness and boundary conditions    339
- 9    Resolvent and uniqueness    340
- 10   Asymptotic stationarity    344

## **Chapter 16     Diffusions**

- 1    The Ornstein-Uhlenbeck process    347
- 2    Processes that are locally Brownian    351
- 3    Brownian motion with boundaries    352
- 4    Feller processes    356
- 5    The natural scale    358
- 6    Speed measure    362
- 7    Boundaries    365
- 8    Construction of Feller processes    370
- 9    The characteristic operator    375
- 10   Uniqueness    379
- 11    $\varphi + (x)$  and  $\varphi - (x)$     383
- 12   Diffusions    385

## **Appendix: On Measure and Function Theory    391**

## **Bibliography    405**

## **Index    412**

*To my mother and father  
and Tuesday's children*

# CHAPTER 1

## INTRODUCTION

A good deal of probability theory consists of the study of limit theorems. These limit theorems come in two categories which we call strong and weak. To illustrate and also to dip into history we begin with a study of coin-tossing and a discussion of the two most famous prototypes of weak and strong limit theorems.

### 1. $n$ INDEPENDENT TOSSES OF A FAIR COIN

These words put us immediately into difficulty. What meaning can be assigned to the words, *coin*, *fair*, *independent*? Take a pragmatic attitude—all computations involving  $n$  tosses of a fair coin are based on two givens:

a) There are  $2^n$  possible outcomes, namely, all sequences  $n$ -long of the two letters  $H$  and  $T$  (Heads and Tails).

b) Each sequence has probability  $2^{-n}$ .

Nothing else is given. All computations regarding odds, and so forth, in fair coin-tossing are based on (a) and (b) above. Hence we take (a) and (b) as being the complete definition of  $n$  independent tosses of a fair coin.

### 2. THE "LAW OF AVERAGES"

Vaguely, almost everyone believes that for large  $n$ , the number of heads is *about the same* as the number of tails. That is, if you toss a fair coin a large number of times, then about half the tosses result in heads.

How to make this mathematics? All we have at our disposal to mathematize the "law of averages" are (a) and (b) above. So if there is anything at all corresponding to the law of averages, it must come out of (a) and (b) with no extra added ingredients.

Analyze the  $2^n$  sequences of  $H$  and  $T$ . In how many of these sequences do exactly  $k$  heads appear? This is a combinatorial problem which clearly can be rephrased as: Given  $n$  squares, in how many different ways can we distribute  $k$  crosses on them? (See Fig. 1.1.) For example, if  $n = 3$ ,  $k = 2$ , then we have the result shown in Fig. 1.2, and the answer is 3.

To get the answer in general, take the  $k$  crosses and subscript them so they become different from each other, that is,  $+_1, +_2, \dots, +_k$ . Now we

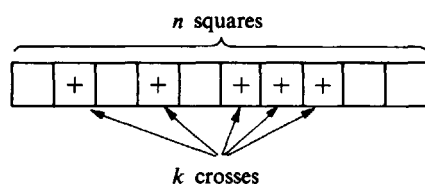


Figure 1.1

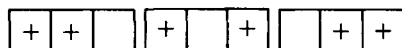


Figure 1.2

may place these latter crosses in  $n$  squares in  $n(n-1)\cdots(n-k+1)$  ways [ $+_1$  may be put down in  $n$  ways, then  $+_2$  in  $(n-1)$  ways, and so forth]. But any permutation of the  $k$  subscripted crosses among the boxes they occupy gives rise to exactly the same distribution of unsubscripted crosses. There are  $k!$  permutations. Hence

**Proposition 1.1.** *There are exactly*

$${}_nC_k = \frac{n!}{k!(n-k)!}$$

*sequences of  $H, T, n$ -long in which  $k$  heads appear.*

Simple computations show that if  $n$  is even,  ${}_nC_k$  is a maximum for  $k = n/2$  and if  $n$  is odd,  ${}_nC_k$  has its maximum value at  $k = (n-1)/2$  and  $k = (n+1)/2$ .

**Stirling's Approximation** [59, Vol. I, pp. 50 ff.]

$$(1.2) \quad n! = e^{-n} n^n \sqrt{2\pi n} (1 + \epsilon_n),$$

where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ .

We use this to get

$$(1.3) \quad {}_{2n}C_n = \frac{(2n)!}{n! n!} = \frac{e^{-2n} (2n)^{2n} \sqrt{4\pi n}}{e^{-2n} n^{2n} (2\pi n)} (1 + \delta_n) \\ = 2^{2n} \cdot \frac{1}{\sqrt{\pi n}} (1 + \delta_n),$$

where  $\delta_n \rightarrow 0$  as  $n \rightarrow \infty$ .

In  $2n$  trials there are  $2^{2n}$  possible sequences of outcomes  $H, T$ . Thus (1.3) implies that  $k = n$  for only a fraction of about  $1/\sqrt{\pi n}$  of the sequences. Equivalently, the probability that the number of heads equals the number of tails is about  $1/\sqrt{\pi n}$  for  $n$  large (see Fig. 1.3).

**Conclusion.** As  $n$  becomes large, the proportion of sequences such that heads comes up exactly  $n/2$  times goes to zero (see Fig. 1.3).

Whatever the "law of averages" may say, it is certainly not reasonable in a thousand tosses of a fair coin to expect exactly 500 heads. It is not

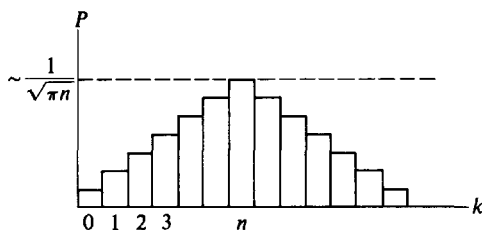


Figure 1.3 Probability of exactly  $k$  heads in  $2n$  tosses.

possible to fix a number  $M$  such that for  $n$  large most of the sequences have the property that the number of heads in the sequence is within  $M$  of  $n/2$ . For  $2n$  tosses this fraction of the sequences is easily seen to be less than  $2M/\sqrt{\pi n}$  (forgetting  $\delta_n$ ) and so becomes smaller and smaller.

To be more reasonable, perhaps the best we can get is that usually the *proportion* of heads in  $n$  tosses is close to  $\frac{1}{2}$ . More precisely—

*Question.* Given any  $\epsilon > 0$ , for how many sequences does the proportion of heads differ from  $\frac{1}{2}$  by less than  $\epsilon$ ?

The answer to this question is one of the earliest and most famous of the limit theorems of probability. Let  $N(n, \epsilon)$  be the number of sequences  $n$ -long satisfying the condition of the above question.

**Theorem 1.4.**  $\lim_n 2^{-n} N(n, \epsilon) = 1$ .

*In other words, the fraction of sequences such that the proportion of heads differs from  $\frac{1}{2}$  by less than  $\epsilon$  goes to one as  $n$  increases for any  $\epsilon > 0$ .*

This theorem is called *the weak law of large numbers for fair coin tossing*. To prove this theorem we need to show that

$$(1.5) \quad \lim_n \left[ \frac{1}{2^n} \sum_{k; |k/n - 1/2| < \epsilon} {}^nC_k \right] = 1.$$

Theorem 1.4 states that most of the time, if you toss a coin  $n$  times, the proportion of heads will be close to  $\frac{1}{2}$ . Is this what is intuitively meant by *the law of averages*? Not quite—the abiding faith seems to be that no matter how badly you have done on the first  $n$  tosses, eventually things will settle down and smooth out *if you keep tossing the coin*.

Ignore this faith for the moment. Let us go back and establish some notation and machinery so we can give Theorem 1.4 an interesting proof. One proof is simply to establish (1.5) by direct computation. It was done this way originally, but the following proof is simpler.



**Definition 1.6**

- a) Let  $\Omega_n$  be the space consisting of all sequences  $n$ -long of  $H, T$ . Denote these sequences by  $\omega_1, \omega_2, \dots, \omega_N, N = 2^n$ .
- b) Let  $A, B, C$ , and so forth, denote subsets of  $\Omega_n$ . The probability  $P(A)$  of any subset  $A$  is defined as the sum of the probabilities of all sequences in  $A$ , that is,

$$P(A) = 2^{-n} \text{ (number of sequences in } A),$$

equivalently,  $P(A)$  is the fraction of the total number of sequences that are in  $A$ .

For example, one interesting subset of  $\Omega_n$  is the set  $A_1$  of all sequences such that the first member is  $H$ . This set can be described as "the first toss results in heads." We should certainly have, if (b) above makes sense,  $P(A_1) = \frac{1}{2}$ . This is so, because there are exactly  $2^{n-1}$  members of  $\Omega_n$  whose first member is  $H$ .

- c) Let  $X(\omega)$  be any real-valued function on  $\Omega_n$ . Define the expected value of  $X$  as

$$EX = \sum_{\omega \in \Omega_n} X(\omega) \cdot \frac{1}{2^n}.$$

Note that the expected value of  $X$  is just its average weighted by the probability. Suppose  $X(\omega)$  takes the value  $x_1$  on the set of sequences  $A_1, A_2$  on  $A_2$ , and so forth; then, of course,

$$EX = \sum_i x_i P(A_i).$$

And also note that  $EX$  is an integral, that is,

$$E(\alpha X + \beta Y) = \alpha EX + \beta EY,$$

where  $\alpha, \beta$  are real numbers, and  $EX \geq 0$ , for  $X \geq 0$ . Also, in the future we will denote by  $\{\omega; \dots\}$  the subset of  $\Omega_n$  satisfying the conditions following the semicolon.

The proof of 1.4 will be based on the important Chebyshev inequality.

**Proposition 1.7.** For  $X(\omega)$  any function on  $\Omega_n$  and any  $\epsilon > 0$ ,

$$P(\omega; |X(\omega)| \geq \epsilon) \leq \frac{1}{\epsilon^2} E(X^2).$$

*Proof*

$$\begin{aligned} P(\omega; |X| \geq \epsilon) &= \frac{1}{2^n} (\text{number of } \omega; |X(\omega)| \geq \epsilon) = \sum_{\{\omega; |X| \geq \epsilon\}} \frac{1}{2^n} \\ &\leq \sum_{\{\omega; |X| \geq \epsilon\}} \frac{X^2(\omega)}{\epsilon^2} \frac{1}{2^n} \leq \frac{1}{\epsilon^2} \sum_{\omega \in \Omega_n} X^2(\omega) \cdot \frac{1}{2^n} = \frac{1}{\epsilon^2} EX^2. \end{aligned}$$

Define functions  $X_1(\omega), \dots, X_n(\omega), S_n(\omega)$  on  $\Omega_n$  by

$$(1.8) \quad X_j(\omega) = \begin{cases} 1 & \text{if } j\text{th member of } \omega \text{ is } H, \\ 0 & \text{if } j\text{th member of } \omega \text{ is } T, \end{cases}$$

$$S_n(\omega) = X_1(\omega) + \dots + X_n(\omega),$$

so that  $S_n(\omega)$  is exactly the number of heads in the sequence  $\omega$ . For practice, note that

$$EX_1 = 0 \cdot P(\omega; \text{first toss} = T) + 1 \cdot P(\omega; \text{first toss} = H) = \frac{1}{2},$$

$$EX_1X_2 = 0 \cdot P(\omega; \text{either first toss or second toss} = T)$$

$$+ 1 \cdot P(\omega; \text{both first toss and second toss} = H) = \frac{1}{4}$$

(since there are  $2^{n-2}$  sequences beginning with  $HH$ ). Similarly, check that if  $i \neq j$ , then

$$(X_i - \frac{1}{2})(X_j - \frac{1}{2}) = \begin{cases} \frac{1}{4} & \text{on } 2^{n-1} \text{ sequences,} \\ -\frac{1}{4} & \text{on } 2^{n-1} \text{ sequences,} \end{cases}$$

so that

$$E(X_i - \frac{1}{2})(X_j - \frac{1}{2}) = 0, \quad i \neq j.$$

Also,

$$(X_i - \frac{1}{2})^2 \equiv \frac{1}{4}, \quad \Rightarrow E(X_i - \frac{1}{2})^2 = \frac{1}{4}.$$

Finally, write

$$S_n - \frac{n}{2} = \sum_{j=1}^n (X_j - \frac{1}{2})$$

so that

$$(1.9) \quad E\left(\frac{S_n}{n} - \frac{1}{2}\right)^2 = \frac{1}{n^2} E\left(\sum_{i,j} (X_i - \frac{1}{2})(X_j - \frac{1}{2})\right) \\ = \frac{1}{n} E(X_1 - \frac{1}{2})^2 = \frac{1}{4n}.$$

*Proof of Theorem 1.4.* By Chebyshev's inequality,

$$P\left(\omega; \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \geq \epsilon\right) \leq \frac{E(S_n/n - \frac{1}{2})^2}{\epsilon^2}.$$

Use (1.9) now to get

$$(1.10) \quad P\left(\omega; \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2},$$

implying

$$\lim_n P\left(\omega; \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \geq \epsilon\right) = 0.$$

Since  $P(\Omega_n) = 1$ , this completes the proof.

**Definition 1.11.** Consider  $n$  independent tosses of a biased coin with probability  $p$  of heads. This is defined by

- a) there are  $2^n$  possible outcomes consisting of all sequences in  $\Omega_n$ .
- b) the probability  $P(\omega)$  of any sequence  $\omega$  is given by

$$P(\omega) = (p^{\# \text{ of } H \text{ in } \omega})(q^{\# \text{ of } T \text{ in } \omega}), \quad q = 1 - p.$$

As before, define  $P(A)$ ,  $A \subset \Omega_n$ , by  $P(A) = \sum_{\omega \in A} P(\omega)$ . For  $X(\omega)$  any real valued function on  $\Omega_n$ , define  $EX = \sum_{\omega \in \Omega_n} X(\omega) P(\omega)$ .

The following problems concern biased coin-tossing.

### Problems

1. Show that  $P(\Omega_n) = 1$ ,  $P(A \cup B) \leq P(A) + P(B)$  with equality if  $A$  and  $B$  are disjoint.
2. Show that

$$P(\omega; S_n(\omega) = k) = {}_nC_k p^k q^{n-k}.$$

3. Show that Chebyshev's inequality 1.7 remains true for biased coin-tossing.
4. Prove the weak law of large numbers in the form: for any  $\epsilon > 0$ ,

$$\lim_n P\left(\omega; \left| \frac{S_n(\omega)}{n} - p \right| \geq \epsilon\right) = 0.$$

5. Using Stirling's approximation, find an approximation to the value of  $P(S_{2n} = n)$ .

**Definition 1.12.** For  $\omega \in \Omega_n$ ,  $\omega = (\omega_1, \dots, \omega_n)$ , where  $\omega_i \in \{H, T\}$ , call  $\omega_i$  the  $i$ th coordinate of  $\omega$  or the outcome of the  $i$ th toss. Any subset  $A \subset \Omega_n$  will be referred to as an event. An event  $A \subset \Omega_n$  will be said to depend only on the  $i_1, \dots, i_k$ th tosses if it is of the form

$$A = \{\omega; (\omega_{i_1}, \dots, \omega_{i_k}) \in E\}, \quad E \subset \Omega_k, \quad k \leq n.$$

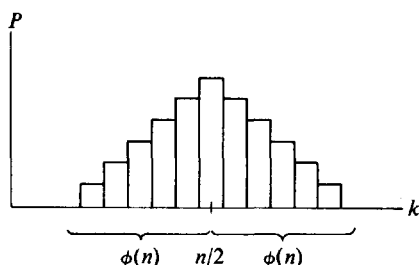
### Problems

6. If  $A$  is of the form above, show that  $P(A) = P'(E)$ , where  $P'(\omega)$  is defined on  $\Omega_k$  by 1.11(b).
7. If  $A, B \subset \Omega_n$  are such that  $A$  depends on the  $i_1, \dots, i_k$  tosses and  $B$  on the  $j_1, \dots, j_m$  tosses and the sets  $(i_1, \dots, i_k), (j_1, \dots, j_m)$  have no common member, then

$$P(A \cap B) = P(A)P(B).$$

[Hint: On Problems 6 and 7 above induction works.]

**Figure 1.4** Probability of  $k$  heads in  $n$  tosses.



### 3. THE BELL-SHAPED CURVE ENTERS (Fluctuation Theory)

For large  $n$ , the weak law of large numbers says that most outcomes have about  $n/2$  heads, more precisely, that the number of heads falls in the range  $(n/2)(1 \pm \epsilon)$  with probability almost one for  $n$  large. Pose the question, how large a fluctuation about  $n/2$  is nonsurprising? For instance, if you get 60  $H$  in 100 tosses, will you strongly suspect the coin of being biased? If you get 54? 43? and so on. Look at the graph in Fig. 1.4. What we want is a function  $\varphi(n)$  increasing with  $n$  such that

$$(1.13) \quad P(|S_n - n/2| < \varphi(n)) \rightarrow \alpha, \quad 0 < \alpha < 1.$$

There are useful things we know about  $\varphi(n)$ . As the maximum height of the graph is order  $1/\sqrt{n}$  we suspect that we will have to go about  $\sqrt{n}$  steps on either side. Certainly if we put  $\varphi(n) = x_n \sqrt{n}/2$  (the  $\frac{1}{2}$  factor to make things work out later on), then  $\lim x_n > 0$ , otherwise the limit in (1.13) would be zero. By Chebyshev's inequality,

$$P(|S_n - n/2| \geq x_n \sqrt{n}/2) \leq n/nx_n^2 = 1/x_n^2.$$

So  $\overline{\lim} x_n < \infty$ , otherwise  $\alpha$  in (1.13) would have to be one. These two bounds lead to the immediate suspicion that we could take  $x_n \rightarrow x$ ,  $0 < x < \infty$ . But there is no reason, then, not to try  $x_n \equiv x$ , all  $n$ . First, examine the case for  $n$  even. We want to evaluate

$$P_n(x) = P(|S_{2n} - n| < x(\sqrt{2n}/2)).$$

This is given by

$$\sum_{|k-n| < x\sqrt{n/2}} 2^{-2n} {}_{2n}C_k.$$

Put  $k = n + j$ , to get

$$P_n(x) = \sum_{|j| < x\sqrt{n/2}} 2^{-2n} {}_{2n}C_{n+j}.$$

$${}_{2n}C_{n+j} = \frac{(2n)!}{(n+j)!(n-j)!},$$

$$(n+j)! = (n+j) \cdots (n+1)n!,$$

$$(n-j)! = n!/n \cdot (n-1) \cdots (n-j+1).$$

Put

$$P_n = P(S_{2n} = n) = 2^{-2n}(2n)!/n! n!$$

and write

$$2^{-2n} {}_{2n}C_{n+j} = P_n \cdot \frac{n(n-1) \cdots (n-j+1)}{(n+j) \cdots (n+1)}.$$

Let  $D_{j,n}$  be the second factor above,

$$D_{j,n} = \frac{1}{(1+j/n)(1+j/(n-1)) \cdots (1+j/(n-j+1))}$$

and

$$\log D_{j,n} = -\sum_{k=0}^{j-1} \log(1+j/(n-k)).$$

Use the expansion  $\log(1+x) = x(1+\epsilon(x))$ , where  $\lim_{x \rightarrow 0} \epsilon(x) = 0$ .

$$\log D_{j,n} = -\sum_{k=0}^{j-1} \frac{j}{n-k} \left(1 + \epsilon\left(\frac{j}{n-k}\right)\right).$$

Note that  $j$  is restricted to the range  $R_n = \{j; |j| < x\sqrt{n/2}\}$ , so that if we write

$$\log D_{j,n} = -(1 + \epsilon_{j,n}) \sum_{k=0}^{j-1} \frac{j}{n-k},$$

then  $\sup_{j \in R_n} \epsilon_{j,n} \rightarrow 0$ . Writing

$$\frac{j}{n-k} = \frac{j}{n} \cdot \frac{1}{1-k/n},$$

since  $0 \leq k < j$ , we find that

$$\log D_{j,n} = -(1 + \epsilon'_{j,n}) \sum_{k=0}^{j-1} \frac{j}{n} = -(1 + \epsilon'_{j,n}) \frac{j^2}{n},$$

where again  $\sup_{j \in R_n} \epsilon'_{j,n} \rightarrow 0$ . Also for  $j \in R_n$ ,  $j^2/n < x^2/2$ , so that

$$D_{j,n} = e^{-j^2/n}(1 + \Delta_{j,n}), \quad j \in R_n,$$

where  $\sup_{j \in R_n} \Delta_{j,n} \rightarrow 0$ . By (1.3),

$$P_n(x) = (1 + \delta_n) \sum_{j \in R_n} \frac{1}{\sqrt{\pi n}} e^{-j^2/n}, \quad \lim_n \delta_n = 0.$$

Make the changes of variable,  $t_j = j\sqrt{2/n}$ ,  $\Delta t = t_{j+1} - t_j = \sqrt{2/n}$ , so the condition  $j \in R_n$  becomes

$$-x < t_j < x.$$

Now the end is near:

$$P_n(x) = (1 + \delta_n) \sum_{-x < t_j < x} \frac{1}{\sqrt{2\pi}} e^{-t_j^2/2} \Delta t.$$

The factor on the right is graciously just the approximating sum for an integral, that is, we have now shown that

$$(1.14) \quad \lim_n P_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{+x} e^{-t^2/2} dt.$$

To get the odd values of  $n$  take  $h > 0$  and note that for  $n$  sufficiently large

$$\begin{aligned} \left\{ \omega; |S_{2n} - n| < \frac{(x-h)}{2} \sqrt{2n} \right\} &\subset \left\{ \omega; \left| S_{2n+1} - \frac{2n+1}{2} \right| < \frac{x}{2} \sqrt{2n+1} \right\} \\ &\subset \left\{ \omega; |S_{2n} - n| < \frac{(x+h)}{2} \sqrt{2n} \right\}, \end{aligned}$$

yielding

$$\begin{aligned} \lim_n P_n(x-h) &\leq \underline{\lim} P \left( \left| S_{2n+1} - \frac{2n+1}{2} \right| < \frac{x}{2} \sqrt{2n+1} \right) \\ &\leq \overline{\lim} P \left( \left| S_{2n+1} - \frac{2n+1}{2} \right| < \frac{x}{2} \sqrt{2n+1} \right) \\ &\leq \lim_n P_n(x+h). \end{aligned}$$

Thus we have proved, as done originally (more or less), a special case of the famous central limit theorem, which along with the law of large numbers shares the throne in probability theory.

### **Theorem 1.15**

$$\lim_n P \left( \left| S_n - \frac{n}{2} \right| < \frac{x}{2} \sqrt{n} \right) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{+x} e^{-t^2/2} dt.$$

There is a more standard form for this theorem: Let

$$(1.16) \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

and  $Z_n = 2S_n - n$ , that is,  $Z_n$  is the excess of heads over tails in  $n$  tosses, or if

$$Y_j = \begin{cases} +1 & \text{if } \omega_j = H, \\ -1 & \text{if } \omega_j = T, \end{cases}$$

then  $Z_n = \sum_1^n Y_j$ . From 1.15

$$\lim_n P \left( \frac{|Z_n|}{\sqrt{n}} < x \right) = \Phi(x) - \Phi(-x).$$

By symmetry,

$$P\left(0 \leq \frac{Z_n}{\sqrt{n}} < x\right) = \frac{1}{2}P\left(\frac{|Z_n|}{\sqrt{n}} < x\right) + \frac{1}{2}P(Z_n = 0),$$

$$P(Z_n < 0) = \frac{1}{2} - \frac{1}{2}P(Z_n = 0),$$

giving

$$\lim_n P\left(\frac{Z_n}{\sqrt{n}} < x\right) = \frac{\Phi(x) + 1 - \Phi(-x)}{2}.$$

But  $\Phi(+\infty) = 1$ , so  $1 - \Phi(-x) = \Phi(x)$ , and therefore,

**Theorem 1.17**

$$\lim_n P\left(\frac{Z_n}{\sqrt{n}} < x\right) = \Phi(x).$$

Thus, the asymptotic distribution of the deviation of the number of heads from  $n/2$  is governed by  $\Phi(x)$ . That  $\Phi(x)$ , the normal curve, should be singled out from among all other limiting distributions is one of the most magical and puzzling results in probability. Why  $\Phi(x)$ ? The above proof gives very little insight as to what properties of  $\Phi(x)$  cause its sudden appearance against the simple backdrop of fair coin-tossing. We return to this later.

### Problems

8. Using Theorem 1.15 as an approximation device, find the smallest integer  $N$  such that in 1600 tosses of a fair coin, there is probability at least 0.99 that the number of heads will fall in the range  $800 \pm N$ .

9. Show that for  $j$  even,

$$P(Z_{2n} = j) = \frac{e^{-j^2/4n}}{\sqrt{\pi n}} (1 + \epsilon_n),$$

where  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ .

10. Show that

$$(1.18) \quad \sup_j |\sqrt{\pi n} P(Z_{2n} = j) - e^{-j^2/4n}| \rightarrow 0,$$

where the sup is over all even  $j$ .

11. Consider a sequence of experiments such that on the  $n$ th experiment a coin is tossed independently  $n$  times with probability of heads  $p_n$ . If  $\lim_n np_n = \lambda$ ,  $0 < \lambda < \infty$ , then letting  $S_n$  be the number of heads occurring in the  $n$ th experiment, show that

$$\lim_n P(S_n = j) = \frac{\lambda^j}{j!} e^{-\lambda}, \quad j = 0, 1, 2, \dots$$

#### 4. STRONG FORM OF THE "LAW OF AVERAGES"

**Definition 1.19.** Let  $\Omega$  be the space consisting of all infinite sequences of  $H$ 's and  $T$ 's. Denote a point in  $\Omega$  by  $\omega$ . Let  $\omega_j$  be the  $j$ th letter in  $\omega$ , that is,  $\omega = (\omega_1, \omega_2, \dots)$ . Define functions on  $\Omega$  as follows:

$$X_j(\omega) = \begin{cases} 1 & \text{if } \omega_j = H, \\ 0 & \text{if } \omega_j = T, \end{cases}$$

$$S_n(\omega) = \sum_{j=1}^n X_j(\omega).$$

We are concerned with the behavior of  $S_n(\omega)/n$  for large  $n$ . The intuitive notion of the law of averages would be that

$$(1.20) \quad \lim_n \frac{S_n(\omega)}{n} = \frac{1}{2}$$

for all  $\omega \in \Omega$ . This is obviously false; the limit need not exist and even if it does, certainly does not need to equal  $\frac{1}{2}$ . [Consider  $\omega = (H, H, H, \dots)$ .] The most we can ask is whether for *almost all* sequences in  $\Omega$ , (1.20) holds.

What about the set  $E$  of  $\omega$  such that (1.20) does not hold? We would like to say that in some sense the probability of this exceptional set is small. Here we encounter an essential difficulty. We know how to assign probabilities to sets of the form

$$(\omega; \omega_1 = H, \dots, \omega_n = T).$$

Indeed, if  $A \subset \Omega_n$ , we know how to assign probabilities to all sets of the form

$$(\omega; (\omega_1, \dots, \omega_n) \in A),$$

that is, simply as before

$$P(\omega; (\omega_1, \dots, \omega_n) \in A) = \frac{1}{2^n} (\text{number of sequences in } A).$$

But the exceptional set  $E$  is the set such that  $S_n(\omega)/n \not\rightarrow \frac{1}{2}$  and so does not depend on  $\omega_1, \dots, \omega_n$  for any  $n$ , but rather on the asymptotic distribution of heads and tails in the sequence  $\omega$ .

But anyhow, let's try to push through a proof and then see what is wrong with it and what needs to be fixed up.

**Theorem 1.21** (Strong law of large numbers). *The probability of the set of sequences  $E$  such that  $S_n(\omega)/n \not\rightarrow \frac{1}{2}$  is zero.*

*Proof.* First note that

$$\lim_n \frac{S_n}{n} = \frac{1}{2} \Leftrightarrow \lim_m \frac{S_m^2}{m^2} = \frac{1}{2},$$



since for any  $n$ , take  $m$  such that

$$m^2 \leq n < (m+1)^2 \quad \text{or} \quad 0 \leq n - m^2 \leq 2m.$$

For this  $m$ ,

$$\begin{aligned} \left| \frac{S_n}{n} - \frac{S_{m^2}}{m^2} \right| &= \left| \frac{S_n}{m^2} - \frac{S_{m^2}}{m^2} + \left( \frac{1}{n} - \frac{1}{m^2} \right) S_n \right| \\ &\leq \frac{|n - m^2|}{m^2} + n \left| \frac{1}{n} - \frac{1}{m^2} \right| \leq \frac{2}{m} + \frac{2}{m} = \frac{4}{m}. \end{aligned}$$

Fix  $\epsilon > 0$  and let  $E_\epsilon$  be the set  $\{\omega; \overline{\lim} |S_{m^2}/m^2 - \frac{1}{2}| > \epsilon\}$ . Look at the set  $E_{m_0, m_1} \subset \Omega$  of sequences such that the inequality  $|S_{m^2}/m^2 - \frac{1}{2}| > \epsilon$  occurs at least once for  $m_0 \leq m \leq m_1$ . That is,

$$E_{m_0, m_1} = \bigcup_{m=m_0}^{m_1} \left\{ \omega; \left| \frac{S_{m^2}}{m^2} - \frac{1}{2} \right| > \epsilon \right\}.$$

The set  $E_{m_0, m_1}$  is a set of sequences that depends only on the coordinates  $\omega_1, \dots, \omega_{m_1^2}$ . We know how to assign probability to such sets, and applying the result of Problem 1,

$$P(E_{m_0, m_1}) \leq \sum_{m=m_0}^{m_1} P\left(\omega; \left| \frac{S_{m^2}}{m^2} - \frac{1}{2} \right| > \epsilon\right).$$

Using Chebyshev's inequality in the form (1.10) we get

$$P(E_{m_0, m_1}) \leq \frac{1}{4\epsilon^2} \sum_{m=m_0}^{m_1} \frac{1}{m^2}.$$

Let  $m_1$  go to infinity and note that

$$E_{m_0} = \lim_{m_1} E_{m_0, m_1} = \bigcup_{m=m_0}^{\infty} \left\{ \omega; \left| \frac{S_{m^2}}{m^2} - \frac{1}{2} \right| > \epsilon \right\}$$

is the set of all sequences such that the inequality  $|S_{m^2}/m^2 - \frac{1}{2}| > \epsilon$  occurs at least once for  $m \geq m_0$ . Also note that the  $\{E_{m_0, m_1}\}$  are an increasing sequence of sets in  $m_1$  for  $m_0$  fixed. *If we could make a vital transition and say that*

$$P(E_{m_0}) = \lim_{m_1} P(E_{m_0, m_1}),$$

then it would follow that

$$P(E_{m_0}) \leq \frac{1}{4\epsilon^2} \sum_{m=m_0}^{\infty} \frac{1}{m^2}.$$

Now  $\overline{\lim} |S_{m^2}/m^2 - \frac{1}{2}| > \epsilon$  if and only if for any  $m_0$ ,  $\exists m > m_0$  such  $|S_{m^2}/m^2 - \frac{1}{2}| > \epsilon$ . From this,  $E_\epsilon = \lim_{m_0} E_{m_0}$ , where the sets  $E_{m_0}$  are decreasing in  $m_0$ . (The limits of increasing or decreasing sequences of sets

are well defined, for example,  $\lim_{m_0} E_{m_0} = \bigcap_{m_0=1}^{\infty} E_{m_0}$ , and so forth.) *If we could again assert as above that*

$$P(E_\epsilon) = \lim_{m_0} P(E_{m_0}),$$

then

$$P(E_\epsilon) \leq \lim_{m_0} \left[ \frac{1}{4\epsilon^2} \sum_{m=m_0}^{\infty} \frac{1}{m^2} \right] = 0.$$

By definition,  $E$  is the set  $\{\omega; \overline{\lim} |S_{m^2}/m^2 - \frac{1}{2}| > 0\}$ , so  $E = \lim_k E_{1/k}$ ,  $k$  running through the positive integers, and the sets  $E_{1/k}$  increasing in  $k$ .

*Once more, if we assert that*

$$P(E) = \lim_k P(E_{1/k}),$$

then since  $P(E_{1/k}) = 0$ , all  $k > 0$ , consequently  $P(E) = 0$  and the theorem is proven. *Q.E.D.???*

The real question is one of how may probability be assigned to subsets of  $\Omega$ . What we need for the above proof is an assignment of probability  $P(\cdot)$  on a class of subsets  $\mathcal{F}$  of  $\Omega$  such that  $\mathcal{F}$  contains all the sets that appear in the above proof and such that  $P(\cdot)$  in some way corresponds to a fair coin-tossing probability. More concretely, what we want are the statements

### (1.22)

- i)  $\mathcal{F}$  contains all subsets depending only on a finite number of tosses, that is, all sets of the form  $\{\omega; (\omega_1, \dots, \omega_n) \in A\}$ ,  $A \subset \Omega_n$ ,  $n \geq 1$ , and  $P(\cdot)$  is defined on these sets by

$$P(\omega; (\omega_1, \dots, \omega_n) \in A) = P_n(A),$$

where  $P_n$  is the probability defined previously on  $\Omega_n$ ;

- ii) if  $A_n$  is any monotone sequence of sets in  $\mathcal{F}$ , then  $\lim_n A_n$  is also in  $\mathcal{F}$ ;  
 iii) if the  $A_n$  are as in (ii) above, then

$$\lim_n P(A_n) = P\left(\lim_n A_n\right);$$

- iv) if  $A, B \in \mathcal{F}$  are disjoint, then

$$P(A \cup B) = P(A) + P(B).$$

Of these four, one is simply the requirement that the assignment be consistent with our previous definition of independent coin-tossing. Two and three are exactly the statement of what is needed to make the transitions in the proof of the law of large numbers valid. Four is that the assignment  $P(\cdot)$  continue to have on  $\Omega$  the property that the probability assignment has on  $\Omega_n$  and whose absence would seem intuitively most offensive, namely,

that if two sets of outcomes are disjoint, then the probability of getting into either one or the other is the sum of the probabilities of each one.

Also, is the assignment of  $P(\cdot)$  unique in any sense? If it is not, then we are in real difficulty. We can put the above questions into more amenable form. Let  $\mathcal{F}_0$  be the class of all subsets of  $\Omega$  depending on only a finite number of tosses, then

**Proposition 1.23.**  $\mathcal{F}_0$  is a field, where

**Definition 1.24.** A class of subsets  $\mathcal{C}$  of a space  $\Omega$  is a field if it is closed under finite unions, intersections, and complementation. The complement of  $\Omega$  is the empty set  $\emptyset$ .

The proof of (1.23) is a direct verification. For economy, take  $\mathcal{F}$  to be the smallest class of sets containing  $\mathcal{F}_0$  such that  $\mathcal{F}$  has property (1.22ii). That such a smallest class exists can be established by considering the sets common to every class of sets containing  $\mathcal{F}_0$  satisfying (1.22ii). But (see Appendix A), these properties imply

**Proposition 1.25.**  $\mathcal{F}$  is the smallest  $\sigma$ -field containing  $\mathcal{F}_0$ , where

**Definition 1.26.** A class of subsets  $\mathcal{F}$  of  $\Omega$  is a  $\sigma$ -field if it is closed under complementation, and countable intersections and unions. For any class  $\mathcal{C}$  of subsets of  $\Omega$ , denote by  $\mathcal{F}(\mathcal{C})$  the smallest  $\sigma$ -field containing  $\mathcal{C}$ .

Also

**Proposition 1.27.**  $P(\cdot)$  on  $\mathcal{F}$  satisfies (1.22) iff  $P(\cdot)$  is a probability measure, where

**Definition 1.28.** A nonnegative set function  $P(\cdot)$  defined on a  $\sigma$ -field  $\mathcal{F}$  of subsets of  $\Omega$  is a probability measure if

- i) (normalization)  $P(\Omega) = 1$ ;
- ii) ( $\sigma$ -additivity) for every finite or countable collection  $\{B_k\}$  of sets in  $\mathcal{F}$  such that  $B_k$  is disjoint from  $B_j$ ,  $k \neq j$ ,

$$P\left(\bigcup_k B_k\right) = \sum_k P(B_k).$$

*Proof of 1.27.* If  $P(\cdot)$  satisfies (1.22), then by finite induction on (iv)

$$P\left(\bigcup_1^n B_k\right) = \sum_1^n P(B_k).$$

Let  $A_n = \bigcup_1^n B_k$ , then the  $A_n$  are a monotone sequence of sets,  $\lim A_n = \bigcup_k B_k$ . By (1.22iii),

$$P\left(\bigcup_k B_k\right) = \lim_n \left(\sum_1^k P(B_k)\right) = \sum_k P(B_k).$$

Conversely, if  $P(\cdot)$  is  $\sigma$ -additive, it implies (1.22). For if the  $\{A_n\}$  are a monotone sequence, say  $A_n \subset A_{n+1}$ , we can let  $B_k = A_k - A_{k-1}$ ,  $k > 1$ ,  $B_1 = A_1$ . The  $\{B_k\}$  are disjoint, and  $\lim_n A_n = \bigcup_k B_k$ . Thus  $\sigma$ -additivity gives

$$P\left(\lim_n A_n\right) = \sum_k P(B_k).$$

Use  $\bigcup_1^n B_k = A_n$ ,  $P(A_n) = \sum_1^n P(B_k)$  to get

$$\lim_n P(A_n) = \lim_n \left( \sum_1^n P(B_k) \right) = \sum_k P(B_k).$$

The starting point is a set function with the following properties:

**Definition 1.29.** A nonnegative set function  $P$  on a field  $\mathcal{F}_0$  is a finite probability measure if

- i)  $P(\Omega) = 1$ ;
- ii) for  $A, B \in \mathcal{F}_0$ , and disjoint,

$$P(A \cup B) = P(A) + P(B).$$

Now the original question can be restated in more standard form: Given the finite probability measure  $P(\cdot)$  defined on  $\mathcal{F}_0$  by (1.22i), does there exist a probability measure defined on  $\mathcal{F}$  and agreeing with  $P(\cdot)$  on  $\mathcal{F}_0$ ? And in what sense is the measure unique? The problem is seen to be one of extension—given  $P(\cdot)$  on  $\mathcal{F}_0$ , is it possible to extend the domain of definition of  $P(\cdot)$  to  $\mathcal{F}$  such that it is  $\sigma$ -additive? But this is a standard measure theoretical question. The surprise is that the attempt to patch up the strong law of large numbers has led directly to this well-known problem (see Appendix A.9).

## 5. AN ANALYTIC MODEL FOR COIN-TOSSING

The fact that the sequence  $X_1(\omega), X_2(\omega), \dots$  comprised functions depending on consecutive independent tosses of a fair coin was to some extent immaterial in the proof of the strong law. For example, produce functions on a different space  $\Omega'$  this way: Toss a well-balanced, six-sided die independently, let  $\Omega'$  be the space of all infinite sequences  $\omega' = (\omega'_1, \omega'_2, \dots)$ , where  $\omega'_k$  takes values in  $(1, \dots, 6)$ . Define  $X'_n(\omega')$  to be one if the  $n$ th throw results in an even face, zero if in an odd face. The sequence  $X'_1(\omega'), X'_2(\omega'), \dots$  has the same probabilistic structure as  $X_1(\omega), \dots$  in the sense that the probability of any sequence  $n$ -long of zeros and ones is  $1/2^n$  in both models (with the appropriate definition of independent throws of a well-balanced die). But this assignment of probabilities is the important information, rather than the exact nature of the underlying space. For example, the same argument

leading to the strong law of large numbers holds for the variables  $X'_1, X'_2, \dots$ . Therefore, in general, we will consider as a model for fair coin-tossing any set of functions  $X_1, X_2, \dots$ , with values zero or one defined on a space  $\Omega$  of points  $\omega$  such that probability  $1/2^n$  is assigned to all sets of the form

$$\{\omega; X_1 = s_1, \dots, X_n = s_n\}$$

for  $s_1, \dots, s_n$  any sequence of zeros and ones.

An interesting analytic model can be constructed on the half-open unit interval  $\Omega = [0, 1)$ . It can be shown that every number  $x$  in  $[0, 1)$  has a unique binary expansion containing an infinite number of zeros. The latter restriction takes care of binary rational points which have two expansions, that is

$$\begin{aligned} \frac{1}{2} &= 0.1000 \dots = \frac{1}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \dots \\ &= 0.0111 \dots = \frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \dots \end{aligned}$$

Now for any  $x \in [0, 1)$  write down this expansion  $x = .x_1x_2\dots$  and define

$$(1.30) \quad X_n(x) = x_n.$$

That is,  $X_n(x)$  is the  $n$ th digit in the expansion of  $x$  (see Fig. 1.5).

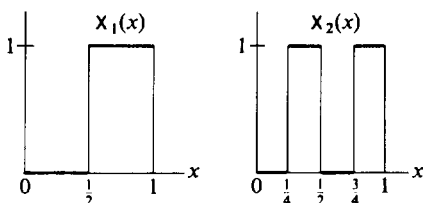


Figure 1.5

To every interval  $I \subset [0, 1)$ , assign probability  $P(I) = \|I\|$ , the length of  $I$ . Now check that the probability of the set

$$\{x; X_1 = s_1, \dots, X_n = s_n\}$$

is  $1/2^n$ , because this set is exactly the interval

$$\left\{x; \frac{s_1}{2} + \frac{s_2}{2^2} + \dots + \frac{s_n}{2^n} \leq x < \frac{s_1}{2} + \frac{s_2}{2^2} + \dots + \frac{s_n}{2^n} + \frac{1}{2^n}\right\}.$$

Thus,  $X_1(x), X_2(x), \dots$  on  $[0, 1)$  with the given probability is a model for fair coin-tossing.

The interest in this particular model is that the extension of  $P$  is a classical result. The smallest  $\sigma$ -field containing all the intervals

$$\{x; X_1(x) = s_1, \dots, X_n(x) = s_n\}$$

is the Borel field  $\mathcal{B}_1([0, 1))$  of subsets of  $[0, 1)$ , and there is a unique extension of  $P$  to a probability on this field, namely, Lebesgue measure. The theorem establishing the existence of Lebesgue measure makes the proof of the strong law of large numbers rigorous for the analytic model. The statement in this context is:

**Theorem 1.31.** *For almost all  $x \in [0, 1)$  with respect to Lebesgue measure, the asymptotic proportions of zero's and one's in the binary expansion of  $x$  is  $\frac{1}{2}$ .*

The existence of Lebesgue measure in the analytic model for coin-tossing makes it plausible that there is an extended probability in the original model. This is true, but we defer the proof to the next chapter.

Another way of looking at the analytic model is to say that the binary expansion of a number in  $[0, 1)$  produces independent zeros and ones with probability  $\frac{1}{2}$  each with respect to Lebesgue measure. Thus, as Theorem 1.31 illustrates, any results established for fair coin-tossing can be written as theorems concerning functions and numbers on and in  $[0, 1)$  with respect to Lebesgue measure. Denote Lebesgue measure from now on by  $dx$  or  $l(dx)$ .

**Problem 12.** (The Rademacher Functions). Let

$$\gamma_i(x) = \begin{cases} -1, & x_i = 1, \\ +1, & x_i = 0, \end{cases}$$

where  $x = .x_1x_2\cdots$  is the unique binary expansion of  $x$  containing an infinite number of zeros. Show from the properties of coin-tossing that if  $i_k \neq i_j$  for  $j \neq k$ ,

$$\int_0^1 \prod_1^m \gamma_{i_k}(x) dx = 0.$$

Graph  $\gamma_3(x)$ ,  $\gamma_4(x)$ . Show that the sequence of functions  $\{\gamma_i(x)\}$  is orthogonal with respect to Lebesgue measure.

## 6. CONCLUSIONS

The strong law of large numbers and the central limit theorem illustrate the two main types of limit theorems in probability.

**Strong limit theorems.** *Given a sequence of functions  $Y_1(\omega), Y_2(\omega), \dots$  there is a limit function  $Y(\omega)$  such that  $P(\omega; \lim_n Y_n(\omega) = Y(\omega)) = 1$ .*

**Weak limit theorems.** *Given a sequence of functions  $Y_1(\omega), Y_2(\omega), \dots$  show that*

$$\lim_n P(\omega; Y_n(\omega) < x)$$

*exists for every  $x$ .*

There is a great difference between strong and weak theorems which will become more apparent. We will show later, for instance, that  $Z_n/\sqrt{n}$  has no limit in any reasonable way. A more dramatic example of this is: on  $([0, 1), \mathcal{B}_1([0, 1)))$  with  $P$  being Lebesgue measure, define

$$Y_n(y) = \begin{cases} 0, & y < \frac{1}{2}, \\ 1, & \frac{1}{2} \leq y < 1, \end{cases}$$

for  $n$  even. For  $n$  odd,

$$Y_n(y) = \begin{cases} 1, & y < \frac{1}{2}, \\ 0, & \frac{1}{2} \leq y < 1. \end{cases}$$

For all  $n$ ,  $P(y; Y_n(y) < x) = P(y; Y_1(y) < x)$ . But for every  $y \in [0, 1)$

$$\varlimsup_n Y_n(y) = 1, \quad \varliminf_n Y_n(y) = 0.$$

To begin with we concentrate on strong limit theorems. But to do this we need a more firmly constructed measure theoretic foundation.

## NOTES

To get some of the fascinating interplay between probability and number theory, refer to Mark Kac's monograph [83].

Although there will be very little subsequent work with combinatorics in this text, they occupy an honored and powerful place in probability theory. First, for many of the more important theorems, the original version was for independent fair coin-tossing. Even outside of this, there are some strong theorems in probability for which the most interesting proofs are combinatorial. A good source for these uses are Feller's books [59].

An elegant approach to the measure theoretic aspects of probability can be found in Neveu's book [113].

## CHAPTER 2

# MATHEMATICAL FRAMEWORK

### 1. INTRODUCTION

The context that is necessary for the strong limit theorems we want to prove is:

**Definition 2.1.** A probability space consists of a triple  $(\Omega, \mathcal{F}, P)$  where

- i)  $\Omega$  is a space of points  $\omega$ , called the sample space and sample points.
- ii)  $\mathcal{F}$  is a  $\sigma$ -field of subsets of  $\Omega$ . These subsets are called events.
- iii)  $P(\cdot)$  is a probability measure on  $\mathcal{F}$ ; henceforth refer to  $P$  as simply a probability.

On  $\Omega$  there is defined a sequence of real-valued functions  $X_1(\omega), X_2(\omega), \dots$  which are random variables in the sense of

**Definition 2.2.** A function  $X(\omega)$  defined on  $\Omega$  is called a random variable if for every Borel set  $B$  in the real line  $R^{(1)}$ , the set  $\{\omega; X(\omega) \in B\}$  is in  $\mathcal{F}$ . ( $X(\omega)$  is a measurable function on  $(\Omega, \mathcal{F})$ .)

Whether a given function is a random variable, of course, depends on the pair  $(\Omega, \mathcal{F})$ . The reason underlying 2.2 is that we want probability assigned to all sets of the form  $\{\omega; X(\omega) \in I\}$ , where  $I$  is some interval. It will follow from 2.29 that if  $\{\omega; X(\omega) \in I\}$  is in  $\mathcal{F}$  for all intervals  $I$ , then  $X$  must be a random variable.

**Definition 2.3.** A countable stochastic process, or process, is a sequence of random variables  $X_1, X_2, \dots$  defined on a common probability space  $(\Omega, \mathcal{F}, P)$ .

But in a probabilistic model arising in gambling or science the given data are usually an assignment of probability to a much smaller class of sets. For example, if all the variables  $X_1, X_2, \dots$  take values in some countable set  $F$ , the probability of all sets of the form

$$(2.4) \quad \{\omega; X_1 = s_1, \dots, X_n = s_n\}, \quad s_k \in F, \quad k = 1, \dots, n$$

is usually given. If the  $X_1, X_2, \dots$  are not discrete, then often the specification is for all sets of the form

$$(2.5) \quad \{\omega; X_1 \in I_1, \dots, X_n \in I_n\},$$

where  $I_1, \dots, I_n$  are intervals.



To justify the use of a probability space as a framework for probability theory it is really necessary to show that a reasonable assignment of probabilities to a small class of sets has a unique extension to a probability  $P$  on a probability space  $(\Omega, \mathcal{F}, P)$ . There are fairly general results to this effect. We defer this until we have explored some of the measure-theoretic properties of processes.

## 2. RANDOM VECTORS

Given two spaces  $\Omega$  and  $R$ , let  $\mathbf{X}$  be a function on  $\Omega$  to  $R$ ,  $\mathbf{X}: \Omega \rightarrow R$ . The inverse image under  $\mathbf{X}$  of a set  $B \subset R$  is  $\{\omega; \mathbf{X} \in B\}$ . We abbreviate this by  $\{\mathbf{X} \in B\}$ .

**Proposition 2.6.** *Set operations are preserved under inverse mappings, that is,*

$$\begin{aligned}\left\{\mathbf{X} \in \bigcup_{\lambda} B_{\lambda}\right\} &= \bigcup_{\lambda} \{\mathbf{X} \in B_{\lambda}\}, \\ \left\{\mathbf{X} \in \bigcap_{\lambda} B_{\lambda}\right\} &= \bigcap_{\lambda} \{\mathbf{X} \in B_{\lambda}\}, \\ \{\mathbf{X} \in B^c\} &= \{\mathbf{X} \in B\}^c.\end{aligned}$$

( $B^c$  denotes the complement of the set  $B$ .)

*Proof.* By definition.

This quickly gives

**Proposition 2.7.** *If  $\mathbf{X}: \Omega \rightarrow R$ , and  $\mathcal{B}$  is a  $\sigma$ -field in  $R$ , the class of sets  $\{\mathbf{X} \in B\}$ ,  $B \in \mathcal{B}$ , is a  $\sigma$ -field. If  $\mathcal{F}$  is a  $\sigma$ -field in  $\Omega$ , then the class of subsets  $B$  in  $R$  such that  $\{\mathbf{X} \in B\} \in \mathcal{F}$  is a  $\sigma$ -field.*

*Proof.* Both assertions are obvious from 2.6.

**Definition 2.8.** *If there are  $\sigma$ -fields  $\mathcal{F}$  and  $\mathcal{B}$ , in  $\Omega$ ,  $R$  respectively,  $\mathbf{X}: \Omega \rightarrow R$  is called a random vector if  $\{\mathbf{X} \in B\} \in \mathcal{F}$ , for all  $B \in \mathcal{B}$ . ( $\mathbf{X}$  is a measurable map from  $(\Omega, \mathcal{F})$  to  $(R, \mathcal{B})$ .)*

We will sometimes refer to  $(R, \mathcal{B})$  as the range space of  $\mathbf{X}$ . But the *range* of  $\mathbf{X}$  is the direct image under  $\mathbf{X}$  of  $\Omega$ , that is, the union of all points  $\mathbf{X}(\omega)$ ,  $\omega \in \Omega$ .

Denote by  $\mathcal{F}(\mathbf{X})$  the  $\sigma$ -field of all sets of the form  $\{\mathbf{X} \in B\}$ ,  $B \in \mathcal{B}$ .

**Definition 2.9.** *If  $\mathcal{A}$  is a  $\sigma$ -field contained in  $\mathcal{F}$ , call  $\mathbf{X}$   $\mathcal{A}$ -measurable if  $\mathcal{F}(\mathbf{X}) \subset \mathcal{A}$ .*

If there is a probability space  $(\Omega, \mathcal{F}, P)$  and  $\mathbf{X}$  is a random vector with range space  $(R, \mathcal{B})$ , then  $\hat{P}$  can be naturally defined on  $\mathcal{B}$  by

$$\hat{P}(B) = P(\mathbf{X} \in B).$$

It is easy to check that  $\hat{P}$  defined this way is a probability on  $\mathcal{B}$ .

**Definition 2.10.**  $\hat{P}$  is called the probability distribution of the random vector  $\mathbf{X}$ .

Conversely, suppose  $\mathbf{X}$  is a random vector on  $(\Omega, \mathcal{F})$  to  $(R, \mathcal{B})$  and there is a probability distribution  $\hat{P}$  defined on  $\mathcal{B}$ . Since every set in  $\mathcal{F}(\mathbf{X})$  is of the form  $\{\mathbf{X} \in B\}$ ,  $B \in \mathcal{B}$ , can  $P$  be defined on  $\mathcal{F}(\mathbf{X})$  by

$$(2.11) \quad P(\mathbf{X} \in B) = \hat{P}(B)?$$

The answer, in general, is no! The difficulty is that the same set  $A \in \mathcal{F}(\mathbf{X})$  may be represented in two different ways as  $\{\mathbf{X} \in B_1\}$  and  $\{\mathbf{X} \in B_2\}$ , and there is no guarantee that  $\hat{P}(B_1) = \hat{P}(B_2)$ . What is true is

**Proposition 2.12.** Let  $F$  be the range of  $\mathbf{X}$ . If  $B \in \mathcal{B}$ ,  $B \subset F^c$  implies  $\hat{P}(B) = 0$ , then  $P$  is uniquely defined on  $\mathcal{F}(\mathbf{X})$  by 2.11, and is a probability.

*Proof.* If  $A = \{\mathbf{X} \in B_1\} = \{\mathbf{X} \in B_2\}$ , then  $B_1 - B_2$  and  $B_2 - B_1$  are both in  $F^c$ . Hence  $\hat{P}(B_1) = \hat{P}(B_2)$ . The  $\sigma$ -additivity is quickly verified.

**Problem 1.** Use 2.12 and the existence of the analytic model for coin-tossing to prove the existence of the desired extension of  $P$  in the original model.

### 3. THE DISTRIBUTION OF PROCESSES

Denote by  $R^{(\infty)}$  the space consisting of all infinite sequences  $(x_1, x_2, \dots)$  of real numbers. In  $R^{(\infty)}$  an  $n$ -dimensional rectangle is a set of the form

$$\{\mathbf{x} \in R^{(\infty)}; x_1 \in I_1, \dots, x_n \in I_n\},$$

where  $I_1, \dots, I_n$  are finite or infinite intervals. Take the Borel field  $\mathcal{B}_\infty$  to be the smallest  $\sigma$ -field of subsets of  $R^{(\infty)}$  containing all finite-dimensional rectangles.

If each component of  $\mathbf{X} = (X_1, \dots)$  is a random variable, then it follows that the vector  $\mathbf{X}$  is a measurable mapping to  $(R^{(\infty)}, \mathcal{B}_\infty)$ . In other words,

**Proposition 2.13.** If  $X_1, X_2, \dots$  are random variables on  $(\Omega, \mathcal{F})$ , then for  $\mathbf{X} = (X_1, X_2, \dots)$  and every  $B \in \mathcal{B}_\infty$ ,  $\{\mathbf{X} \in B\} \in \mathcal{F}$ .

*Proof.* Let  $S$  be a finite-dimensional rectangle

$$S = \{\mathbf{x} \in R^{(\infty)}; x_1 \in I_1, \dots, x_n \in I_n\}.$$

Then

$$\{\mathbf{X} \in S\} = \{X_1 \in I_1\} \cap \{X_2 \in I_2\} \cap \dots \cap \{X_n \in I_n\}.$$

This is certainly in  $\mathcal{F}$ . Now let  $\mathcal{C}$  be the class of sets  $C$  in  $\mathcal{B}_\infty$  such that  $\{\mathbf{X} \in C\} \in \mathcal{F}$ . By 2.7  $\mathcal{C}$  is a  $\sigma$ -field. Since  $\mathcal{C}$  contains all rectangles,  $\mathcal{C} = \mathcal{B}_\infty$ .

If all that we observe are the values of a process  $X_1(\omega), X_2(\omega), \dots$  the underlying probability space is certainly not uniquely determined. As an example, suppose that in one room a fair coin is being tossed independently,

and calls zero or one are being made for tails or heads respectively. In another room a well-balanced die is being cast independently and zero or one called as the resulting face is odd or even. There is, however, no way of discriminating between these two experiments on the basis of the calls.

Denote  $\mathbf{X} = (X_1, \dots)$ . From an observational point of view, the thing that really interests us is not the space  $(\Omega, \mathcal{F}, P)$ , but the distribution of the values of  $\mathbf{X}$ . If two processes,  $\mathbf{X}$  on  $(\Omega, \mathcal{F}, P)$ ,  $\mathbf{X}'$  on  $(\Omega', \mathcal{F}', P')$  have the same probability distribution,

$$(2.14) \quad \hat{P}(B) = P(\mathbf{X} \in B) = P'(\mathbf{X}' \in B), \quad \text{all } B \in \mathcal{B}_\infty,$$

then there is no way of distinguishing between the processes by observing them.

**Definition 2.15.** Two processes  $\{\mathbf{X}_n\}$  on  $(\Omega, \mathcal{F}, P)$  and  $\{\mathbf{X}'_n\}$  on  $(\Omega', \mathcal{F}', P')$  will be said to have the same distribution if (2.14) holds.

The distribution of a process contains all the information which is relevant to probability theory. All theorems we will prove depend only on the distribution of the process, and hence hold for all processes having that distribution. Among all processes having a given distribution  $\hat{P}$  on  $\mathcal{B}_\infty$ , there is one which has some claim to being the simplest.

**Definition 2.16.** For any given distribution  $\hat{P}$  define random variables  $\hat{X}_1, \hat{X}_2, \dots$  on  $(R^{(\infty)}, \mathcal{B}_\infty, \hat{P})$  by

$$\hat{X}_n(x_1, x_2, \dots) = x_n.$$

This process is called the coordinate representation process and has the same distribution as the original process.

This last assertion is immediate since for any  $B \in \mathcal{B}_\infty$ ,

$$\hat{P}(\mathbf{x}; (\hat{X}_1, \hat{X}_2, \dots) \in B) = \hat{P}(\mathbf{x}; \mathbf{x} \in B) = \hat{P}(B).$$

This construction also leads to the observation that given any probability  $\hat{P}$  on  $\mathcal{B}_\infty$ , there exists a process  $\mathbf{X}$  such that  $P(\mathbf{X} \in B) = \hat{P}(B)$ .

Define the Borel field  $\mathcal{B}_n$  in  $R^{(n)}$  as the smallest  $\sigma$ -field containing all rectangles  $\{(x_1, \dots, x_n); x_1 \in I_1, \dots, x_n \in I_n\}$ ,  $I_1, \dots, I_n$  intervals.

**Definition 2.17.** An  $n$ -dimensional cylinder set in  $R^{(\infty)}$  is any set of the form  $\{\mathbf{x} \in R^{(\infty)}; (x_1, \dots, x_n) \in B\}$ ,  $B \in \mathcal{B}_n$ .

## Problems

2. Show that the class of all finite-dimensional cylinder sets is a field, but not a  $\sigma$ -field.

3. Let  $F$  be a countable set,  $F = \{r_j\}$ . Denote by  $F^{(\infty)}$  the set of all infinite sequences with coordinates in  $F$ . Show that  $\mathbf{X}: \Omega \rightarrow F$  is a random variable with respect to  $(\Omega, \mathcal{F})$  iff  $\{\omega; \mathbf{X}(\omega) = r_j\} \in \mathcal{F}$ , all  $j$ .

4. Given two processes  $X, X'$  such that both take values in  $F^{(\infty)}$ , show that they have the same distribution iff

$$P(X_1 = t_1, \dots, X_n = t_n) = P'(X'_1 = t_1, \dots, X'_n = t_n)$$

for every  $n$ -sequence  $(t_1, \dots, t_n)$ ,  $t_k \in \{r_j\}$ ,  $k = 1, \dots, n$ .

#### 4. EXTENSION IN SEQUENCE SPACE

Given the concept of the distribution of a process, the extension problem can be looked at in a different way. The given data is a specification of values  $P(X \in B)$  for a class of sets in  $\mathcal{B}_\infty$ . That is, a set function  $\hat{P}$  is defined for a class of sets  $\mathcal{C} \subset \mathcal{B}_\infty$ , and  $\hat{P}(B)$  is the probability that the observed values of the process fall in  $B \in \mathcal{C}$ . Now ask: Does there exist *any* process whose distribution agrees with  $\hat{P}$  on  $\mathcal{C}$ ? Alternatively—construct a process  $X$  such that  $P(X \in B) = \hat{P}(B)$  for all  $B \in \mathcal{C}$ . This is equivalent to the question of whether  $\hat{P}$  on  $\mathcal{C}$  can be extended to a probability on  $\mathcal{B}_\infty$ . Because if so, the coordinate representation process has the desired distribution. As far as the original sample space is concerned, once  $\hat{P}$  on  $\mathcal{B}_\infty$  is gotten, 2.12 can be used to get an extension of  $P$  to  $\mathcal{F}(X)$ , if  $\hat{P}$  assigns probability zero to sets  $B \in \mathcal{B}_\infty$  falling in the complement of the range of  $X$ .

Besides this, another reason for looking at the extension problem on  $(R^{(\infty)}, \mathcal{B}_\infty)$  is that this is the smoothest space on which we can always put a process having any given distribution. It has some topological properties which allow nice extension results to be proved.

The basic extension theorem we use is the analog in  $\mathcal{B}_\infty$  of the extension of measures on the real line from their values on intervals. Let  $\mathcal{C}$  be the class of all finite-dimensional rectangles, and assume that  $\hat{P}$  is defined on  $\mathcal{C}$ . A finite-dimensional rectangle may be written as a disjoint union of finite-dimensional rectangles, for instance,

$$\{\mathbf{x} \in R^{(\infty)}; x_1 \in [0, 1]\} = \{\mathbf{x} \in R^{(\infty)}; x_1 \in [0, \tfrac{1}{2}]\} \cup \{\mathbf{x} \in R^{(\infty)}; x_1 \in [\tfrac{1}{2}, 1]\}.$$

Of course, we will insist that if a rectangle  $S$  is a finite union  $\bigcup_j S_j$  of disjoint rectangles, then  $\hat{P}(S) = \sum_j \hat{P}(S_j)$ . But an additional regularity condition is required, simply because the class of finite probabilities is much larger than the class of probabilities.

**Extension Theorem 2.18.** *Let  $\hat{P}$  be defined on the class  $\mathcal{C}$  of all finite-dimensional rectangles and have the properties:*

- $\hat{P}(\cdot) \geq 0$ ,  $\hat{P}(R^{(\infty)}) = 1$ ;
- if  $\{S_j\}$ ,  $j = 1, \dots, m$ , are disjoint  $n$ -dimensional rectangles and

$$S = \bigcup_{j=1}^m S_j$$

is a rectangle, then

$$\hat{P}(S) = \sum_{j=1}^m \hat{P}(S_j);$$

c) if  $\{S_j\}$  are a nondecreasing sequence of  $n$ -dimensional rectangles, and  $S_j \uparrow S$ ,

$$\lim_j \hat{P}(S_j) = \hat{P}(S).$$

Then there is a unique extension of  $\hat{P}$  to a probability on  $\mathcal{B}_\infty$ .

*Proof.* As this result belongs largely to the realm of measure theory rather than probability, we relegate its proof to Appendix A.48.

Theorem 2.18 translates into probability language as: If probability is assigned in a reasonable way to rectangles, then there exists a process  $X_1, X_2, \dots$  such that  $P(X_1 \in I, \dots, X_n \in I_n)$  has the specified values.

If the probabilities are assigned to rectangles, then in order to be well defined, the assignment must be *consistent*. This means here that since an  $n$ -dimensional rectangle is also an  $(n+1)$ -dimensional rectangle (take  $I_{n+1} = R^{(1)}$ ), its assignment as an  $(n+1)$ -dimensional rectangle must agree with its assignment as an  $n$ -dimensional rectangle.

Now consider the situation in which the probability distributions of all finite collections of random variables in a process are specified. Specifically, probabilities  $\hat{P}_n$  on  $\mathcal{B}_n$ ,  $n = 1, 2, \dots$ , are given and  $\hat{P}$  is defined on the class of all finite-dimensional cylinder sets (2.17) by

$$\hat{P}(\mathbf{x} \in R^{(\infty)}; (x_1, \dots, x_n) \in B) = \hat{P}_n(B), \quad B \in \mathcal{B}_n.$$

In order for  $\hat{P}$  to be well-defined, the  $\hat{P}_n$  must be consistent—every  $n$ -dimensional cylinder set is also an  $(n+1)$ -dimensional cylinder set and must be given the same probability by  $\hat{P}_n$  and  $\hat{P}_{n+1}$ .

**Corollary 2.19.** (*Kolmogorov extension theorem*). *There is a unique extension of  $\hat{P}$  to a probability on  $\mathcal{B}_\infty$ .*

*Proof.*  $\hat{P}$  is defined on the class  $\mathcal{C}$  of all finite-dimensional rectangles and is certainly a finite probability on  $\mathcal{C}$ . Let  $S^*, S_j^*$  be rectangles in  $R^{(n)}$ ,  $S_j^* \uparrow S^*$ , then

$$\hat{P}(\mathbf{x} \in R^{(\infty)}; (x_1, \dots, x_n) \in S_j^*) = \hat{P}_n(S_j^*).$$

Since  $\hat{P}_n$  is a probability on  $\mathcal{B}_n$ , it is well behaved under monotone limits (1.27). Hence  $\hat{P}_n(S_j^*) \uparrow \hat{P}_n(S^*)$ , and Theorem 2.18 is in force.

The extension requirements become particularly simple when the required process takes only values in a countable set.

**Corollary 2.20.** *Let  $F \subset R^{(1)}$  be a countable set. If  $p(s_1, \dots, s_n)$  is specified for all finite sequences of elements of  $F$  and satisfies*

$$\begin{aligned} p(\cdot) &\geq 0, \\ \sum_{s_{n+1} \in F} p(s_1, \dots, s_{n+1}) &= p(s_1, \dots, s_n), \\ \sum_{s_1 \in F} p(s_1) &= 1, \end{aligned}$$

then there exists a process  $X_1, X_2, \dots$  such that

$$P(X_1 = s_1, \dots, X_n = s_n) = p(s_1, \dots, s_n).$$

*Proof.* Let  $s$  denote an  $n$ -tuple with coordinates in  $F$ . For any  $B \in \mathcal{B}_n$ , define

$$\hat{P}_n(B) = \sum_{s \in B} p(s).$$

It is easy to check that  $\hat{P}_n$  is a finite probability on  $\mathcal{B}_n$ . Furthermore,  $B_k \uparrow B$  implies

$$\lim_k \hat{P}_n(B_k) = \hat{P}_n(B).$$

Thus we conclude (see 1.27) that  $\hat{P}_n$  is a probability on  $\mathcal{B}_n$ . The  $\hat{P}_n$  are clearly consistent. Now apply 2.19 to get the result.

The extension results in this section are a bit disquieting, because even though the results are purely measure-theoretic, the proofs in the space  $(R^{(\infty)}, \mathcal{B}_\infty)$  depend essentially on the topological properties of Euclidean spaces. This is in the nature of the problem. For example, if one has an infinite product  $(\Omega^{(\infty)}, \mathcal{F}_\infty)$  of  $(\Omega, \mathcal{F})$  spaces, that is:

$$\Omega^{(\infty)} = \{(\omega_1, \dots); \omega_k \in \Omega\};$$

$\mathcal{F}_\infty$  the smallest  $\sigma$ -field containing all sets of the form

$$\{(\omega_1, \dots); \omega_1 \in A_1, \dots, \omega_n \in A_n\}, \quad A_1, \dots, A_n \in \mathcal{F};$$

$P_n$  a probability on all  $n$ -dimensional cylinder sets; and the set  $\{P_n\}$  consistent; then a probability  $P$  on  $\mathcal{F}_\infty$  agreeing with  $P_n$  on  $n$ -dimensional cylinder sets may or may not exist. For a counter-example, see Jessen and Andersen [77] or Neveu [113, p. 84].

**Problem 5.** Take  $\Omega = (0, 1]$ ,

$$X_n(x) = \begin{cases} 1, & 0 < x < 1/n, \\ 0, & 1/n < x \leq 1. \end{cases}$$

Let  $\mathcal{F}_0 = \bigcup_{n \geq 1} \mathcal{F}(X_1, \dots, X_n)$ . Characterize the sets in  $\mathcal{F}_0$ ,  $\mathcal{F}(\mathcal{F}_0)$ . For  $B \in \mathcal{B}_n$ , assign

$$P((X_1, \dots, X_n) \in B) = \begin{cases} 1, & \text{if } (1, 1, \dots, 1) \in B, \\ 0, & \text{otherwise.} \end{cases}$$

Prove that  $P$  is additive on  $\mathcal{F}_0$ , but that there is no extension of  $P$  to a probability on  $\mathcal{F}(\mathcal{F}_0)$ .

## 5. DISTRIBUTION FUNCTIONS

What is needed to ensure that two processes have the same distribution?

**Definition 2.21.** Given a process  $\{X_n\}$  on  $(\Omega, \mathcal{F}, P)$ , define the  $n$ -dimensional distribution functions by

$$F_n(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n).$$

The functions  $F_n(\cdot)$  are real-valued functions defined on  $R^{(n)}$ . Denote these at times by  $F_n(\mathbf{x}_n)$  or, to make their dependence on the random variables explicit, by  $F_{X_1 \dots X_n}(x_1, \dots, x_n)$  or  $F_{\mathbf{X}_n}(\mathbf{x}_n)$ .

**Theorem 2.22.** *Two processes have the same distribution iff all their distribution functions are equal.*

The proof of 2.22 follows from a more general result that we want on the record.

**Proposition 2.23.** *Let  $Q, Q'$  be two probabilities on  $(\Omega, \mathcal{F})$ . Let  $\mathcal{C}$  be a class of sets such that  $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$ , and  $\mathcal{F} = \mathcal{F}(\mathcal{C})$ . Then  $Q = Q'$  on  $\mathcal{C}$  implies that  $Q = Q'$  on  $\mathcal{F}$ .*

There seems to be a common belief that 2.23 is true without the hypothesis that  $\mathcal{C}$  be closed under  $\cap$ . To disprove this, let  $\Omega = \{a, b, c, d\}$ ,  $Q_1(a) = Q_1(d) = Q_2(b) = Q_2(c) = \frac{1}{8}$ , and  $Q_1(b) = Q_1(c) = Q_2(a) = Q_2(d) = \frac{3}{8}$ .  $\mathcal{F}$  is the class of all subsets of  $\Omega$ , and

$$\mathcal{C} = \{a \cup b, d \cup c, a \cup c, b \cup d\}.$$

*Proof.* Let  $\mathcal{F}_0(\mathcal{C})$  be the smallest field containing  $\mathcal{C}$ . By the unique extension theorem it suffices to show that  $Q = Q'$  on  $\mathcal{F}_0(\mathcal{C})$ . Let  $\mathcal{D}$  be the smallest class of sets such that

- i)  $\mathcal{C} \subset \mathcal{D}$ ,
- ii)  $\Omega \in \mathcal{D}$ ,
- iii)  $A, B \in \mathcal{D}, B \subset A \Rightarrow A - B \in \mathcal{D}$ .

Then  $\mathcal{D} = \mathcal{F}_0(\mathcal{C})$ . To see this, let  $\mathcal{U}$  be the class of sets  $A$  in  $\mathcal{D}$  such that  $A \cap C \in \mathcal{D}$  for all  $C \in \mathcal{C}$ . Then notice that  $\mathcal{U}$  satisfies (i), (ii), (iii) above, so  $\mathcal{U} = \mathcal{D}$ . This implies that  $A \cap C \in \mathcal{D}$  for all  $A \in \mathcal{D}, C \in \mathcal{C}$ . Now let  $\mathcal{E}$  be the class of sets  $E$  in  $\mathcal{D}$  such that  $A \cap E \in \mathcal{D}$ , all  $A \in \mathcal{D}$ . Similarly  $\mathcal{E}$  satisfies (i), (ii), (iii), so  $\mathcal{E} = \mathcal{D}$ . This yields  $\mathcal{D}$  closed under  $\cap$ , but by (ii), (iii),  $\mathcal{D}$  is also closed under complementation, proving the assertion. Let  $\mathcal{G}$  be the class of sets  $G$  in  $\mathcal{F}$  such that  $Q(G) = Q'(G)$ . Then  $\mathcal{G}$  satisfies (i), (ii), (iii)  $\Rightarrow \mathcal{D} \subset \mathcal{G}$  or  $\mathcal{F}_0(\mathcal{C}) \subset \mathcal{G}$ .

*Returning to the proof of 2.22.* Let  $\hat{P}, \hat{P}'$  be defined on  $\mathcal{B}_\infty$  by  $P(\mathbf{X} \in B)$ ,  $P'(\mathbf{X}' \in B)$ , respectively. Let  $\mathcal{C} \subset \mathcal{B}_\infty$  be the class of all sets of the form  $C = \{\mathbf{x}; x_1 < y_1, \dots, x_n < y_n\}$ . Then clearly  $\mathcal{C}$  is closed under  $\cap$ , and  $\mathcal{F}(\mathcal{C}) = \mathcal{B}_\infty$ . Now  $\hat{P}(C) = F_{\mathbf{X}_n}(y_1, \dots, y_n)$  and  $\hat{P}'(C) = F_{\mathbf{X}'_n}(y_1, \dots, y_n)$ , so that  $\hat{P} = \hat{P}'$  on  $\mathcal{C}$  by hypothesis. By 2.23  $\hat{P} = \hat{P}'$  on  $\mathcal{B}_\infty$ .

Another proof of 2.22 which makes it more transparent is as follows: For any function  $G(x_1, \dots, x_n)$  on  $R^{(n)}$  and  $I$  an interval  $[a, b]$ ,  $a < b$ ,  $\mathbf{x} = (x_1, \dots, x_n)$ , write

$$\begin{aligned} \underline{\hat{K}}_1 G(\mathbf{x}) &= G(x_1, \dots, x_{k-1}, b, x_{k+1}, \dots, x_n) \\ &\quad - G(x_1, \dots, x_{k-1}, a, x_{k+1}, \dots, x_n). \end{aligned}$$

By definition, since

$$F_n(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n),$$

the probability of any rectangle  $\{X_1 \in I_1, \dots, X_n \in I_n\}$  with  $I_1, \dots, I_n$  left closed, right open, can be expressed in terms of  $F_n$  by

$$(2.24) \quad P(X_1 \in I_1, \dots, X_n \in I_n) = \hat{\Delta}_{I_1} \cdots \hat{\Delta}_{I_n} F_n(\mathbf{x})$$

because, for  $I_n = [a_n, b_n)$ ,

$$\begin{aligned} \hat{\Delta}_{I_n} F_n(\mathbf{x}) &= P(X_1 < x_1, \dots, X_n < b_n) - P(X_1 < x_1, \dots, X_n < a_n) \\ &= P(X_1 < x_1, \dots, X_{n-1} < x_{n-1}, X_n \in I_n). \end{aligned}$$

By taking limits, we can now get the probabilities of all rectangles. From the extension theorem 2.18 we know that specifying  $P$  on rectangles uniquely determines it.

Frequently, the distribution of a process is specified by giving a set of distribution functions  $\{F_n(\mathbf{x})\}$ ,  $n = 1, 2, \dots$ . But in order that  $\{F_n(\mathbf{x})\}$  be derived from a process  $\{X_n\}$  on a probability space  $(\Omega, \mathcal{F}, P)$ , they must have certain essential properties.

**Proposition 2.25.** *The distribution functions  $F_n(\mathbf{x})$  satisfy the conditions:*

i) *Non-negativity.* For finite intervals  $I_k = [a_k, b_k)$ ,  $k = 1, \dots, n$ ,

$$\hat{\Delta}_{I_1} \cdots \hat{\Delta}_{I_n} F_n(\mathbf{x}) \geq 0.$$

ii) *Continuity from below.* If  $\mathbf{x}^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$  and  $x_j^{(k)} \uparrow x_j$ ,  $j = 1, \dots, n$ , then

$$F_n(\mathbf{x}^{(k)}) \uparrow F_n(\mathbf{x}).$$

iii) *Normalization.* All limits of  $F_n$  exist as

$$x_j \uparrow +\infty \quad \text{or} \quad x_j \downarrow -\infty.$$

If  $x_j \downarrow -\infty$ , then  $F_n(\mathbf{x}) \rightarrow 0$ . If all  $x_j$ ,  $j = 1, \dots, n \uparrow +\infty$ , then  $F_n(\mathbf{x}) \rightarrow 1$ .

*The set of distribution functions are connected by*

iv) *Consistency*

$$\lim_{x_n \uparrow +\infty} F_n(x_1, \dots, x_n) = F_{n-1}(x_1, \dots, x_{n-1}), \quad n > 1.$$

*Proof.* The proof of (i) follows from (2.24). To prove (ii), note that

$$\{\omega; X_1 < x_1^{(k)}, \dots, X_n < x_n^{(k)}\} \uparrow \{\omega; X_1 < x_1, \dots, X_n < x_n\}.$$

Use the essential fact that probabilities behave nicely under monotone limits to get (ii). Use this same fact to prove (iii) and (iv); e.g. if  $x_j \downarrow -\infty$ , then  $\{\omega; X_1 < x_1, \dots, X_n < x_n\} \downarrow \emptyset$ . If all  $x_1, \dots, x_n \uparrow +\infty$ , then

$$\{\omega; X_1 < x_1, \dots, X_n < x_n\} \uparrow \Omega.$$



Another important construction theorem verifies that the conditions of 2.25 characterize the distribution functions of a process.

**Theorem 2.26.** *Given a set of functions  $\{F_n(\mathbf{x})\}$  satisfying 2.25 (i), (ii), (iii), (iv), there is a process  $\{X_n\}$  on  $(\Omega, \mathcal{F}, P)$  such that*

$$P(X_1 < x_1, \dots, X_n < x_n) = F_n(x_1, \dots, x_n).$$

*Proof.* The idea of how the proof should go is simple. Use  $\Omega = R^{(\infty)}$ ,  $\mathcal{F} = \mathcal{B}_\infty$ , and use the coordinate representation process  $\hat{X}_1, \hat{X}_2, \dots$ . We want to construct  $\hat{P}$  on  $\mathcal{B}_\infty$  such that if  $S \in \mathcal{B}_\infty$  is a semi-infinite rectangle of the form

$$\{\mathbf{x} \in R^{(\infty)}; x_1 < y_1, \dots, x_n < y_n\},$$

then

$$\hat{P}(S) = F_n(y_1, \dots, y_n).$$

To construct  $\hat{P}$  starting from  $F_n$ , define  $\hat{P}$  on rectangles whose sides are left closed, right open, intervals  $I_1, \dots, I_n$  by

$$\hat{P}(\mathbf{x}; x_1 \in I_1, \dots, x_n \in I_n) = \hat{\Delta}_{I_1} \cdots \hat{\Delta}_{I_n} F_n(\mathbf{x}).$$

Extend this to all rectangles by taking limits. The consistency 2.25 (iv) guarantees that  $\hat{P}$  is well defined on all rectangles. All that is necessary to do now is to verify the conditions of 2.18. If  $S_j, S$  are left closed, right open rectangles, and  $S_j \uparrow S$ , then the continuity from below of  $F_n$ , 2.25 (ii), yields

$$\lim_j \hat{P}(S_j) = \hat{P}(S).$$

To verify the above for general rectangles, use the fact that their probabilities can be defined as limits of probabilities of left closed, right open rectangles.

The complication is in showing additivity of  $\hat{P}$  on rectangles. It is sufficient to show that

$$\hat{P}(S) = \sum_j \hat{P}(S_j)$$

for left closed, right open, disjoint rectangles  $S_1, \dots, S_k$  whose union is a rectangle  $S$ . In one dimension the statement  $\hat{P}(S) = \sum_j \hat{P}(S_j)$  follows from the obvious fact that for  $a_0 < a_1 < \dots < a_k$ ,

$$\Delta_{[a_0, a_k]} F_1(x_1) = \sum_{j=1}^k \Delta_{[a_{j-1}, a_j]} F_1(x_1).$$

The general result is a standard theorem in the theory of the Stieltjes integral (McShane [111a, pp. 245–246]).

If a function  $F(x_1, \dots, x_n)$  satisfies only the first three conditions of 2.25 then Theorem 2.26 implies the following.

**Corollary 2.27.** *There are random variables  $X_1, \dots, X_n$  on a space  $(\Omega, \mathcal{F}, P)$  such that*

$$P(X_1 < x_1, \dots, X_n < x_n) = F(x_1, \dots, x_n).$$

*Hence, any such function will be called an  $n$ -dimensional distribution function.*

If a set  $\{F_n\}$ ,  $n = 1, 2, \dots$ , of  $n$ -dimensional distribution functions satisfies 2.25 (iv), call them *consistent*. The specification of a consistent set of  $\{F_n\}$  is pretty much the minimum amount of data needed to completely specify the distribution of a process in the general case.

### Problems

6. For any random variable  $X$ , let  $F_X(x) = P(X < x)$ . The function  $F_X(x)$  is called the distribution function of the variable  $X$ .

Prove that  $F_X(x)$  satisfies

- i)  $x \leq x' \Rightarrow F_X(x) \leq F_X(x')$ ,
- ii)  $\lim_{x \uparrow +\infty} F_X(x) = 1, \quad \lim_{x \downarrow -\infty} F_X(x) = 0$ ,
- iii)  $\lim_{x_k \uparrow x} F_X(x_k) = F_X(x)$ .

7. If a function  $F(x_1, \dots, x_n)$  is nondecreasing in each variable separately, does this imply  $\bigwedge_{I_1} \cdots \bigwedge_{I_n} F_n(\mathbf{x}) > 0$ ? Give an example of a function  $F(x, y)$  such that

- i)  $F(x, y) \geq 0$ ,
- ii)  $x' > x, y' > y \Rightarrow F(x', y) \geq F(x, y), F(x, y') \geq F(x, y)$ .
- iii) There are finite intervals  $I_1, I_2$  such that

$$\bigwedge_{I_1} \bigwedge_{I_2} F(x, y) < 0.$$

8. Let  $F_1(x), F_2(x), \dots$  be functions satisfying the conditions of Problem 6. Prove that the functions

$$F_n(x_1, \dots, x_n) = \prod_{k=1}^n F_k(x_k)$$

form a consistent set of distribution functions.

### 6. RANDOM VARIABLES

From now on, for reasons sufficient and necessary, we study random variables defined on a probability space. The sufficient reason is that the extension theorems state that given a fairly reasonable assignment of probabilities, a process can be constructed fitting the specified data. The necessity is that most strong limit theorems require this kind of an environment. Now we record a few facts regarding random variables and probability spaces.

**Proposition 2.28.** Let  $\mathcal{C}$  be a class of Borel sets such that  $\mathcal{F}(\mathcal{C}) = \mathcal{B}_1$ ,  $X$  a real-valued function on  $\Omega$ . If  $\{X \in C\} \in \mathcal{F}$ , all  $C \in \mathcal{C}$ , then  $X$  is a random variable on  $(\Omega, \mathcal{F})$ .

*Proof.* Let  $\mathcal{D} \subset \mathcal{B}_1$  be the class of all Borel sets  $D$  such that  $\{X \in D\} \in \mathcal{F}$ .  $\mathcal{D}$  is a  $\sigma$ -field.  $\mathcal{C} \subset \mathcal{D} \Rightarrow \mathcal{D} = \mathcal{B}_1$ .

**Corollary 2.29.** If  $\{X \in I\} \in \mathcal{F}$  for all intervals  $I$ , then  $X$  is a random variable.

At times functions come up which may be infinite on some parts of  $\Omega$  but which are random variables on subsets where they are finite.

**Definition 2.30.** An extended random variable  $X$  on  $(\Omega, \mathcal{F})$  may assume the values  $\pm \infty$ , but  $\{X \in B\} \in \mathcal{F}$ , for all  $B \in \mathcal{B}_1$ .

**Proposition 2.31.** Let  $X$  be a random vector to  $(R, \mathcal{B})$ . If  $\varphi(x)$  is a random variable on  $(R, \mathcal{B})$ , then  $\varphi(X)$  is a random variable on  $(\Omega, \mathcal{F})$ , measurable  $\mathcal{F}(X)$ .

*Proof.* Write, for  $A \in \mathcal{B}_1$ ,

$$\{\varphi(X) \in A\} = \{X \in \varphi^{-1}(A)\} \in \mathcal{F}(X),$$

$\varphi^{-1}(A)$  here denoting the inverse image of  $A$  under  $\varphi$ .

**Definition 2.32.** For random variables  $X_1, X_2, \dots$  on  $\Omega$ , the  $\sigma$ -field of all events depending on the first  $n$  outcomes is the class of sets  $\{(X_1, \dots, X_n) \in B\}$ ,  $B \in \mathcal{B}_n$ . Denote this by  $\mathcal{F}(X_1, \dots, X_n)$ . The class of sets depending on only a finite number of outcomes is

$$\mathcal{F}_0 = \bigcup_{n \geq 1} \mathcal{F}(X_1, \dots, X_n).$$

In general,  $\mathcal{F}_0$  is a field, but not a  $\sigma$ -field. But the fact that  $\mathcal{F}(X) = \mathcal{F}(\mathcal{F}_0)$  follows immediately from the definitions.

**Proposition 2.33.** Given a process  $X_1, X_2, \dots$ . For every set  $A_1 \in \mathcal{F}(X)$  and  $\epsilon > 0$ , there is a set  $A_2$  in some  $\mathcal{F}(X_1, \dots, X_n)$  such that

$$P(A_1 \triangle A_2) \leq \epsilon.$$

( $A_1 \triangle A_2$  is the symmetric set difference  $(A_1 - A_2) \cup (A_2 - A_1)$ .)

*Proof.* The proof of this is one of the standard results which cluster around the construction used in the Carathéodory extension theorem. The statement is that if  $P$  on  $\mathcal{F}(\mathcal{F}_0)$  is an extension of  $P$  on  $\mathcal{F}_0$ , then for every set  $A_1 \in \mathcal{F}(\mathcal{F}_0)$  and  $\epsilon > 0$ , there is a set  $A_2$  in the field  $\mathcal{F}_0$  such that  $P(A_2 \triangle A_1) \leq \epsilon$  (see Appendix A.12). Then 2.33 follows because  $\mathcal{F}(X)$  is the smallest  $\sigma$ -field containing

$$\bigcup_{n \geq 1} \mathcal{F}(X_1, \dots, X_n).$$

If all the random variables in a process  $X_1, X_2, \dots$  take values in a Borel set  $E \in \mathcal{B}_1$ , it may be more convenient to use the range space  $(E^{(\infty)}, \mathcal{B}_\infty(E))$ , where  $\mathcal{B}_\infty(E)$  consists of all sets in  $\mathcal{B}_\infty$  which are subsets of  $E^{(\infty)}$ . For example, if  $X_1, X_2, \dots$  are coin-tossing variables, then each one takes values in  $\{0, 1\}$ , and the relevant  $R, \mathcal{B}$  for the process is

$$(\{0, 1\}^{(\infty)}, \mathcal{B}_\infty(\{0, 1\})).$$

If a random variable  $X$  has distribution function  $F(x)$ , then  $P(X \in B)$  is a probability measure on  $\mathcal{B}_1$  which is an extension of the measure on intervals  $[a, b]$  given by  $F(b) - F(a)$ . Thus, use the notation:

**Definition 2.34.** For  $X$  a random variable, denote by  $P(X \in dx)$  or  $F(dx)$  the probability measure  $P(X \in B)$  on  $\mathcal{B}_1$ . Refer to  $F(dx)$  as the distribution of  $X$ .

**Definition 2.35.** A sequence  $X_1, X_2, \dots$  of random variables all having the same distribution  $F(dx)$  are called identically distributed. Similarly, call random vectors  $X_1, X_2, \dots$  with the same range space  $(R, \mathcal{B})$  identically distributed if they have the common distribution

$$\hat{P}(dx) = P(X_n \in dx).$$

### Problems

9. Show that  $\mathcal{B}_\infty(\{0, 1\})$  is the smallest  $\sigma$ -field containing all sets of the form

$$\{x \in \{0, 1\}^{(\infty)}; x_1 = s_1, \dots, x_n = s_n\},$$

where  $s_1, \dots, s_n$  is any sequence  $n$ -long of zeros and ones,  $n = 1, 2, \dots$

10. Given a process  $X_1, X_2, \dots$  on  $(\Omega, \mathcal{F}, P)$ . Let  $m_1, m_2, \dots$  be positive integer-valued random variables on  $(\Omega, \mathcal{F}, P)$ . Prove that the sequence  $X_{m_1}, X_{m_2}, \dots$  is a process on  $(\Omega, \mathcal{F}, P)$ .

## 7. EXPECTATIONS OF RANDOM VARIABLES

**Definition 2.36.** Let  $X$  be a random variable on  $(\Omega, \mathcal{F}, P)$ . Define the expectation of  $X$ , denoted  $EX$ , by  $\int X(\omega) dP(\omega)$ . This is well defined if  $E|X| < \infty$ . Alternative notations for the integrals are

$$\int X dP, \quad \int X(\omega) P(d\omega).$$

**Definition 2.37.** For any probability space  $(\Omega, \mathcal{F}, P)$  define

i) if  $A \in \mathcal{F}$ , the set indicator  $\chi_A(\omega)$  is the random variable

$$\chi_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in A^c. \end{cases}$$

ii) If  $X$  is a random variable, then  $X^+$ ,  $X^-$  are the random variables

$$X^+(\omega) = \begin{cases} X(\omega), & \omega \in \{X \geq 0\}, \\ 0, & \text{otherwise;} \end{cases}$$

$$X^-(\omega) = \begin{cases} -X(\omega), & \omega \in \{X \leq 0\}, \\ 0, & \text{otherwise.} \end{cases}$$

A number of results we prove in this and later sections depend on a principle we state as

**Proposition 2.38.** Consider a class  $\mathcal{L}$  of random variables having the properties

- i)  $X, Y \in \mathcal{L}, \alpha, \beta \geq 0 \Rightarrow \alpha X + \beta Y \in \mathcal{L}$ .
- ii)  $X_n \in \mathcal{L}, X_n(\omega) \uparrow X(\omega)$  for every  $\omega \Rightarrow X \in \mathcal{L}$ .
- iii) For every set  $A \in \mathcal{F}, \chi_A(\omega) \in \mathcal{L}$ .

Then  $\mathcal{L}$  includes all nonnegative random variables on  $(\Omega, \mathcal{F}, P)$ .

*Proof.* See Appendix A.22.

This is used to prove

**Proposition 2.39.** Let the processes  $X$  on  $(\Omega, \mathcal{F}, P)$ ,  $X'$  on  $(\Omega', \mathcal{F}', P')$  have the same distribution. Then if  $\varphi(x)$  is measurable  $(R^{(\infty)}, \mathcal{B}_\infty)$ ,

$$(2.40) \quad \int_{\Omega} \varphi(X(\omega)) P(d\omega) = \int_{\Omega'} \varphi(X'(\omega')) P'(d\omega'),$$

in the sense that if either side is well defined, so is the other, and the two are equal.

*Proof.* Consider all  $\varphi$  for which (2.40) is true. This class satisfies (i) and (ii) of 2.38. Further, let  $B \in \mathcal{B}_\infty$  and let  $\varphi(x) = \chi_B(x)$ . Then the two sides of (2.40) become  $P(X \in B)$  and  $P'(X' \in B)$ , respectively. But these are equal since the processes have the same distribution. Hence (2.40) holds for all nonnegative  $\varphi$ . Thus, for any  $\varphi$ , it holds true for  $|\varphi|$ ,  $\varphi^+$ ,  $\varphi^-$ .

**Corollary 2.41.** Define  $\hat{P}(\cdot)$  on  $\mathcal{B}_\infty$  by  $\hat{P}(B) = P(X \in B)$ . Then if  $\varphi$  is measurable  $(R^{(\infty)}, \mathcal{B}_\infty)$  and  $E|\varphi(X)| < \infty$ ,

$$\int \varphi(X) dP = \int \varphi(x) \hat{P}(dx).$$

*Proof.*  $\{X_n\}$  on  $(\Omega, \mathcal{F}, P)$  and  $\{\hat{X}_n\}$  on  $(R^{(\infty)}, \mathcal{B}_\infty, \hat{P})$ , have the same distribution, where  $\hat{X}_n$  is the coordinate representation process. Thus

$$\int \varphi(X_1, X_2, \dots) dP = \int \varphi(\hat{X}_1, \hat{X}_2, \dots) \hat{P}(d\mathbf{x}),$$

but  $\hat{X}_n(\mathbf{x}) = x_n$ .

## 8. CONVERGENCE OF RANDOM VARIABLES

Given a sequence of random variables  $\{X_n\}$ , there are various modes of strong convergence of  $X_n$  to a limiting random variable  $X$ .

### Definition 2.42

i)  $X_n$  converges to  $X$  almost surely (a.s.) if

$$P(\omega; \lim X_n(\omega) = X(\omega)) = 1.$$

Denote this by  $X_n \xrightarrow{\text{a.s.}} X$  or  $X_n \rightarrow X$  a.s.

ii)  $X_n$  converges to  $X$  in  $r$ th mean, for  $r > 0$ , if  $E|X_n - X|^r \rightarrow 0$ . Denote this by  $X_n \xrightarrow{r} X$ .

iii)  $X_n$  converges in probability to  $X$  if for every  $\epsilon > 0$ ,

$$P(|X_n - X| > \epsilon) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Denote this by  $X_n \xrightarrow{P} X$ .

The important things to notice are:

First, all these convergences are “probabilistic.” That is, if  $X, X_1, \dots$  has the same distribution as  $X', X'_1, \dots$ , then  $X_n \rightarrow X$  in any of the above senses implies that  $X'_n \rightarrow X'$  in the same sense. This is obvious for  $\xrightarrow{r}, \xrightarrow{P}$ . See Problem 12 for  $\xrightarrow{\text{a.s.}}$ .

Secondly, Cauchy convergence in any one of these senses gives convergence.

**Proposition 2.43.** *If  $X_m - X_n \rightarrow 0$  in any of the above ways as  $m, n \rightarrow \infty$  in any way, then there is a random variable  $X$  such that  $X_n \rightarrow X$  in the same way.*

*Proof.* Do a.s. convergence first. For all  $\omega$  such that  $X_n(\omega)$  is Cauchy convergent,  $\lim_n X_n(\omega)$  exists. Hence  $P(\omega; \lim_n X_n(\omega) \text{ exists}) = 1$ . Let  $X(\omega) = \lim_n X_n(\omega)$  for all  $\omega$  such that the limit exists, otherwise put it equal to zero, then  $X_n \xrightarrow{\text{a.s.}} X$ . For the other modes of Cauchy convergence the proof is deferred until Section 3 of the next chapter (Problems 6 and 7).

Thirdly, of these various kinds of convergences  $\xrightarrow{\text{a.s.}}$  is usually the hardest to establish and more or less the strongest. To get from a.s. convergence to  $\xrightarrow{r}$ , some sort of boundedness condition is necessary. Recall

**Theorem 2.44.** (Lebesgue bounded convergence theorem). *If  $Y_n \xrightarrow{\text{a.s.}} Y$  and if there is a random variable  $Z \geq 0$  such that  $EZ < \infty$ , and  $|Y_n| \leq Z$  for all  $n$ , then  $EY_n \rightarrow EY$ . (See Appendix A.28).*

Hence, using  $Y_n = |X_n - X|^r$  in 2.44, we get

$$(2.45) \quad X_n \xrightarrow{\text{a.s.}} X, |X_n|^r \leq Z, EZ < \infty \Rightarrow X_n \xrightarrow{r} X.$$

Convergence in probability is the weakest. The implications go

$$(2.46) \quad \begin{array}{l} \text{i) } \xrightarrow{\text{a.s.}} \Rightarrow \xrightarrow{P} \\ \text{ii) } \xrightarrow{r} \Rightarrow \xrightarrow{P}. \end{array}$$

### Problems

11. Prove (2.46i and ii). [Use a generalization of Chebyshev's inequality on (ii).]

12. Let  $\{X_n\}$ ,  $\{X'_n\}$  have the same distribution. Prove that if  $X_n \rightarrow X$  a.s., there is a random variable  $X'$  such that  $X'_n \rightarrow X'$  a.s.

13. For a process  $\{X_n\}$  prove that the set

$$\left\{ \omega; \lim_n X_n(\omega) \text{ does not exist} \right\}$$

is an event (i.e., is in  $\mathcal{F}$ ).

14. Prove that for  $X$  a random variable with  $E|X| < \infty$ , then  $A_n \in \mathcal{F}$ ,  $A_n \downarrow \emptyset$ , implies

$$\lim_n \int_{A_n} X dP = 0.$$

### NOTES

The use of a probability space  $(\Omega, \mathcal{F}, P)$  as a context for probability theory was formalized by Kolmogorov [98], in a monograph published in 1933. But, as Kolmogorov pointed out, the concept had already been current for some time.

Subsequent work in probability theory has proceeded, almost without exception, from this framework. There has been controversy about the correspondence between the axioms for a probability space and more primitive intuitive notions of probability. A different approach in which the probability of an event is defined as its asymptotic frequency is given by von Mises [112]. The argument can go on at several levels. At the top is the contention that although it seems reasonable to assume that  $P$  is a finite probability, there are no strong intuitive grounds for assuming it  $\sigma$ -additive. Thus, in their recent book [40] Dubins and Savage assume only finite additivity, and even within this weaker framework prove interesting limit theorems. One level down is the question of whether a probability measure  $P$  need be additive at all. The more basic property is argued to be:  $A \subset B \Rightarrow P(A) \leq P(B)$ . But, as always, with weaker assumptions fewer nontrivial theorems can be proven.

At the other end, it happens that some  $\sigma$ -fields have so many sets in them that examples occur which disturb one's intuitive concept of probability. Thus, there has been some work in the direction of restricting the type of  $\sigma$ -field to be considered. An interesting article on this is by Blackwell [8].

For a more detailed treatment of measure theory than is possible in Appendix A, we recommend the books of Neveu [113], Loève [108], and Halmos [64].



## CHAPTER 3

### INDEPENDENCE

Independence, or some form of it, is one of the central concepts of probability, and it is largely responsible for the distinctive character of probability theory.

#### 1. BASIC DEFINITIONS AND RESULTS

##### *Definition 3.1*

- (a) Given random variables  $X_1, \dots, X_n$ , on  $(\Omega, \mathcal{F}, P)$ , they are said to be independent if for any sets  $B_1, \dots, B_n \in \mathcal{B}_1$ ,

$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{k=1}^n P(X_k \in B_k).$$

- (b) Given a probability space  $(\Omega, \mathcal{F}, P)$  and  $\sigma$ -fields  $\mathcal{F}_1, \dots, \mathcal{F}_n$  contained in  $\mathcal{F}$ , they are said to be independent if for any sets  $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$

$$P(A_1 \cap \dots \cap A_n) = \prod_{k=1}^n P(A_k).$$

Obviously,  $X_1, \dots, X_n$  are independent random variables iff  $\mathcal{F}(X_1), \dots, \mathcal{F}(X_n)$  are independent  $\sigma$ -fields.

These definitions have immediate generalizations to random vectors.

**Definition 3.2.** Random vectors  $X_1, X_2, \dots, X_n$  are said to be independent if the  $\sigma$ -fields  $\mathcal{F}(X_1), \mathcal{F}(X_2), \dots, \mathcal{F}(X_n)$ , are independent.

Virtually all the results of this section stated for independent random variables hold for independent random vectors. But as the generalization is so apparent, we usually omit it.

**Definition 3.3.** The random variables  $X_1, X_2, \dots$  are called independent if for every  $n \geq 2$ , the random variables  $X_1, X_2, \dots, X_n$ , are independent.

**Proposition 3.4.** Let  $X_1, X_2, \dots$  be independent random variables and  $B_1, B_2, \dots$  any sets in  $\mathcal{B}_1$ . Then

$$P(X_1 \in B_1, X_2 \in B_2, \dots) = \prod_{k=1}^{\infty} P(X_k \in B_k).$$

*Proof.* Let  $A_n = \{X_1 \in B_1, \dots, X_n \in B_n\}$ . Then  $A_n$  are decreasing, hence

$$P(\lim A_n) = \lim P(A_n).$$

But,

$$\lim A_n = \{X_1 \in B_1, X_2 \in B_2, \dots\}$$

and

$$\lim P(A_n) = \lim_n \prod_1^n P(X_k \in B_k) = \prod_1^\infty P(X_k \in B_k).$$

Note that the same result holds for independent  $\sigma$ -fields.

**Proposition 3.5.** *Let  $X_1, X_2, \dots$  be independent random variables,  $(i_1, i_2, \dots)$ ,  $(j_1, j_2, \dots)$  disjoint sets of integers. Then the fields*

$$\mathcal{F}_1 = \mathcal{F}(X_{i_1}, X_{i_2}, \dots), \mathcal{F}_2 = \mathcal{F}(X_{j_1}, X_{j_2}, \dots)$$

*are independent.*

*Proof.* Consider any set  $D \in \mathcal{F}_2$  of the form  $D = \{X_{j_1} \in B_1, \dots, X_{j_m} \in B_m\}$ ,  $B_k \in \mathcal{B}_1, k = 1, \dots, m$ . Define two measures  $Q_1$  and  $Q'_1$  on  $\mathcal{F}_1$  by, for  $A \in \mathcal{F}_1$ ,

$$Q_1(A) = P(A \cap D), \quad Q'_1(A) = P(A)P(D).$$

Consider the class of sets  $\mathcal{C} \subset \mathcal{F}_1$  of the form

$$C = \{X_{i_1} \in E_1, \dots, X_{i_n} \in E_n\}, E_l \in \mathcal{B}_1, l = 1, \dots, n.$$

Note that

$$\begin{aligned} Q_1(C) &= P\left(\bigcap_{l=1}^n \{X_{i_l} \in E_l\} \bigcap_{k=1}^m \{X_{j_k} \in B_k\}\right) \\ &= \prod_{l=1}^n P(X_{i_l} \in E_l) \prod_{k=1}^m P(X_{j_k} \in B_k) \\ &= P(C)P(D) = Q'_1(C). \end{aligned}$$

Thus  $Q_1 = Q'_1$  on  $\mathcal{C}$ ,  $\mathcal{C}$  is closed under  $\cap$ ,  $\mathcal{F}(\mathcal{C}) = \mathcal{F}_1 \Rightarrow Q_1 = Q'_1$  on  $\mathcal{F}_1$  (see 2.23). Now repeat the argument. Fix  $A \in \mathcal{F}_1$  and define  $Q_2(\cdot)$ ,  $Q'_2(\cdot)$  on  $\mathcal{F}_2$  by  $P(A \cap \cdot)$ ,  $P(A)P(\cdot)$ . By the preceding, for any  $D$  of the form given above,  $Q_2(D) = Q'_2(D)$ , implying  $Q_2 = Q'_2$  on  $\mathcal{F}_2$  and thus for any  $A_1 \in \mathcal{F}_1$ ,  $A_2 \in \mathcal{F}_2$ ,

$$P(A_1 \cap A_2) = P(A_1)P(A_2).$$

**Corollary 3.6.** *Let  $X_1, X_2, \dots$  be independent random variables,  $J_1, J_2, \dots$  disjoint sets of integers. Then the  $\sigma$ -fields  $\mathcal{F}_k = \mathcal{F}(\{X_j\}, j \in J_k)$  are independent.*

*Proof.* Assume that  $\mathcal{F}_1, \dots, \mathcal{F}_n$  are independent. Let

$$J = \bigcup_1^n J_k, \quad \text{and} \quad \mathcal{F}' = \mathcal{F}(\{X_j\}, j \in J).$$

Since  $J$  and  $J_{n+1}$  are disjoint,  $\mathcal{F}'$  and  $\mathcal{F}_{n+1}$  satisfy the conditions of Proposition 3.5 above. Let  $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$ . Then since  $\mathcal{F}_k \subset \mathcal{F}'$ ,  $k = 1, \dots, n$ ,

$$A_1 \cap \dots \cap A_n \in \mathcal{F}'.$$

Let  $A' = A_1 \cap \dots \cap A_n$  and  $A_{n+1} \in \mathcal{F}_{n+1}$ . By 3.5,

$$P(A_{n+1} \cap A') = P(A_{n+1})P(A') = P(A_{n+1})P(A_n) \cdots P(A_1)$$

by the induction hypothesis.

From 3.5 and 3.6 we extract more concrete and interesting consequences. For instance, 3.5 implies that the fields  $\mathcal{F}(X_1, \dots, X_n)$  and  $\mathcal{F}(X_{n+1}, \dots)$  are independent. As another example, if  $\varphi_1, \varphi_2, \dots$  are measurable ( $R^{(m)}, \mathcal{B}_m$ ), then the random variables

$$Z_1 = \varphi_1(X_1, \dots, X_m), \quad Z_2 = \varphi_2(X_{m+1}, \dots, X_{2m}), \dots$$

are independent. Another way of stating 3.6 is to say that the random vectors  $\mathbf{X}_k = (\{X_j\}, j \in J_k)$ , are independent.

How and when do we get independent random variables?

**Theorem 3.7.** *A necessary and sufficient condition for  $X_1, X_2, \dots$ , to be independent random variables is that for every  $n$ , and  $n$ -tuple  $(x_1, \dots, x_n)$*

$$F_{X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n).$$

*Proof.* It is obviously necessary—consider the sets  $\{X_1 < x_1, \dots, X_n < x_n\}$ . To go the other way, we want to show that for arbitrary  $B_1, \dots, B_n \in \mathcal{B}_1$ ,

$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_1^n P(X_k \in B_k).$$

Fix  $x_2, \dots, x_n$  and define two  $\sigma$ -additive measures  $Q$  and  $Q'$  on  $\mathcal{B}_1$  by

$$Q_1(B) = P(X_1 \in B, X_2 < x_2, \dots, X_n < x_n),$$

$$Q'_1(B) = P(X_1 \in B)P(X_2 < x_2, \dots, X_n < x_n).$$

Now on all sets of the form  $(-\infty, x)$ ,  $Q$  and  $Q'$  agree, implying that  $Q = Q'$  on  $\mathcal{B}_1$ . Repeat this by fixing  $B_1 \in \mathcal{B}_1, x_3, \dots, x_n$  and defining

$$Q_2(B) = P(X_1 \in B_1, X_2 \in B, X_3 < x_3, \dots, X_n < x_n),$$

$$Q'_2(B) = P(X_1 \in B_1)P(X_2 \in B)P(X_3 < x_3, \dots, X_n < x_n),$$

so  $Q_2 = Q'_2$  on the sets  $(-\infty, x)$ , hence  $Q_2 = Q'_2$  on  $\mathcal{B}_1$ , and continue on down.

The implication of this theorem is that if we have any one-dimensional distribution functions  $F_1(x), F_2(x), \dots$  and we form the consistent set of distribution functions (see Problem 8, Chapter 2)  $F_1(x_1) \cdots F_n(x_n)$ , then

any resulting process  $X_1, X_2, \dots$  having these distribution functions consists of independent random variables.

**Proposition 3.8.** *Let  $X$  and  $Y$  be independent random variables,  $f$  and  $g$   $\mathcal{B}_1$ -measurable functions such that  $E|f(X)| < \infty$ ,  $E|g(Y)| < \infty$ , then*

$$E|f(X)g(Y)| < \infty$$

and

$$E[f(X)g(Y)] = [Ef(X)][Eg(Y)].$$

*Proof.* For any set  $A \in \mathcal{B}_1$ , take  $f(x) = \chi_A(x)$ ; and consider the class  $\mathcal{L}$  of nonnegative functions  $g(y)$  for which the equality in 3.8 holds.  $\mathcal{L}$  is closed under linear combinations. By the Lebesgue monotone convergence theorem applied to both sides of the equation in 3.8,  $g_n \in \mathcal{L}$ ,  $g_n \uparrow g \Rightarrow g \in \mathcal{L}$ . If  $B \in \mathcal{B}_1$  and  $g(y) = \chi_B(y)$ , then the equation becomes

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B),$$

which holds by independence of  $X$  and  $Y$ . By 2.38,  $\mathcal{L}$  includes all nonnegative  $g$ -measurable  $\mathcal{B}_1$ . Now fix  $g$  and apply to  $f$  to conclude that 3.8 is valid for all nonnegative  $g$  and  $f$ .

For general  $g$  and  $f$ , note that 3.8 holding for nonnegative  $g$  and  $f$  implies  $E|f(X)g(Y)| = [E|f(X)|][E|g(Y)|]$ , so integrability of  $f(X)$  and  $g(Y)$  implies that of  $f(X)g(Y)$ . By writing  $f = f^+ - f^-$ ,  $g = g^+ - g^-$  we obtain the general result.

Note that if  $X$  and  $Y$  are independent, then so are the random variables  $f(X)$  and  $g(Y)$ . So actually the above proposition is no more general than the statement: *Let  $X$  and  $Y$  be independent random variables,  $E|X| < \infty$ ,  $E|Y| < \infty$ , then  $E|XY| < \infty$  and  $EXY = EX \cdot EY$ .*

By induction, we get

**Corollary 3.9.** *Let  $X_1, \dots, X_n$  be independent random variables such that  $E|X_k| < \infty$ ,  $k = 1, \dots, n$ . Then*

$$E|X_1 \cdots X_n| < \infty \quad \text{and} \quad E(X_1 \cdots X_n) = \prod_{k=1}^n EX_k.$$

*Proof.* Follows from 3.8 by induction.

## Problems

1. Let  $\mathcal{F}_1, \mathcal{F}_2$  be independent  $\sigma$ -fields. Show that if a set  $A$  is both in  $\mathcal{F}_1$  and  $\mathcal{F}_2$ , then  $P(A) = 0$  or 1.
2. Use Fubini's theorem (Appendix A.37) to show that for  $X$  and  $Y$  independent random variables
  - a) for any  $B \in \mathcal{B}_1$ ,  $P(X \in B - y)$  is a  $\mathcal{B}_1$ -measurable function of  $y$ ,
  - b)  $P(X + Y \in B) = \int P(X \in B - y)P(Y \in dy)$ .

## 2. TAIL EVENTS AND THE KOLMOGOROV ZERO-ONE LAW

Consider the set  $E$  again, on which  $S_n/n \rightarrow \frac{1}{2}$  for fair coin-tossing. As pointed out, this set has the odd property that whether or not  $\omega \in E$  does not depend on the first  $n$  coordinates of  $\omega$  no matter how large  $n$  is. Sets which have this fascinating property we call tail events.

**Definition 3.10.** Let  $X_1, X_2, \dots$  be any process. A set  $E \in \mathcal{F}(X)$  will be called a tail event if  $E \in \mathcal{F}(X_n, X_{n+1}, \dots)$ , all  $n$ . Equivalently, let  $\mathcal{J}$  be the  $\sigma$ -field  $\bigcap_{n=1}^{\infty} \mathcal{F}(X_n, X_{n+1}, \dots)$ , then  $\mathcal{J}$  is called the tail  $\sigma$ -field and any set  $E \in \mathcal{J}$  is called a tail event.

This definition may seem formidable, but it captures formally the sense in which certain events do not depend on any finite number of their coordinates. For example,

$$E = \left\{ \omega; \frac{X_1 + \dots + X_n}{n} \rightarrow \frac{1}{2} \right\}$$

is a tail event. Because for any  $k \geq 1$ ,

$$E = \left\{ \omega; \frac{X_k + \dots + X_n}{n} \rightarrow \frac{1}{2} \right\},$$

hence  $E \in \mathcal{F}(X_k, X_{k+1}, \dots)$  for all  $k \geq 1$ ,  $\Rightarrow E \in \mathcal{J}$ .

An important class of tail events is given as follows:

**Definition 3.11.** Let  $X_1, X_2, \dots$  be any process,  $B_1, B_2, \dots$  Borel sets. The set  $X_n$  in  $B_n$  infinitely often, denoted  $\{X_n \in B_n \text{ i.o.}\}$  is the set  $\{\omega; \text{ }_n X(\omega) \in B_n \text{ occurs for an infinite number of } n\}$ . Equivalently,

$$\{X_n \in B_n \text{ i.o.}\} = \lim_{m \uparrow \infty} \bigcup_{n=m}^{\infty} \{X_n \in B_n\}.$$

It is fairly apparent that for many strong limit theorems the events involved will be tail. Hence it is most gratifying that the following theorem is in force.

**Theorem 3.12.** (Kolmogorov zero-one law). Let  $X_1, X_2, \dots$  be independent random variables. Then if  $E \in \mathcal{J}$ ,  $P(E)$  is either zero or one.

*Proof.*  $E \in \mathcal{F}(X)$ . By 2.33, there are sets  $E_n \in \mathcal{F}(X_1, \dots, X_n)$  such that  $P(E_n \triangle E) \rightarrow 0$ . This implies  $P(E_n) \rightarrow P(E)$ , and  $P(E_n \cap E) \rightarrow P(E)$ . But  $E \in \mathcal{F}(X_{n+1}, X_{n+2}, \dots)$ , hence  $E$  and  $E_n$  are in independent  $\sigma$ -fields. Thus  $P(E_n \cap E) = P(E_n)P(E)$ . Taking limits in this latter equation gives

$$P(E) = [P(E)]^2.$$

The only solutions of  $x = x^2$  are  $x = 0$  or  $1$ . Q.E.D.

This is really a heart-warming result. It puts us into secure business with strong limit theorems for independent random variables involving tail events. Either the theorem holds true for almost all  $\omega \in \Omega$  or it fails almost surely.

### Problems

3. Show that  $\{X_n \in B_n \text{ i.o.}\}$  is a tail event.
4. In the coin-tossing game let  $s$  be any sequence  $m$ -long of zeros or ones. Let  $Z_n$  be the vector  $(X_{n+1}, \dots, X_{n+m})$ , and  $F$  the set  $\{Z_n = s \text{ i.o.}\}$ . Show that  $F \in \mathcal{F}$ .
5. (the random signs problem). Let  $c_n$  be any sequence of real numbers. In the fair coin-tossing game let  $Y_n = \pm 1$  as the  $n$ th toss is  $H$  or  $T$ . Let  $D = \{\omega; \sum c_n Y_n \text{ converges}\}$ ; show that  $D \in \mathcal{F}$ .

### 3. THE BOREL-CANTELLI LEMMA

Every tail event has probability zero or one. Now the important question is: how to decide which is which. The Borel-Cantelli lemma is a most important step in that direction. It applies to a class of events which includes many tail-events, but it also has other interesting applications.

**Definition 3.13.** In  $(\Omega, \mathcal{F}, P)$ , let  $A_n \in \mathcal{F}$ . The set  $\{A_n \text{ i.o.}\}$  is defined as  $\{\omega; \omega \in A_n \text{ for an infinite number of } n\}$ , or equivalently

$$\{A_n \text{ i.o.}\} = \lim_{m \rightarrow \infty} \bigcup_{n > m} A_n.$$

#### Borel-Cantelli Lemma 3.14

I. The direct half. If  $A_n \in \mathcal{F}$ , then  $\sum_1^\infty P(A_n) < \infty$  implies  $P(A_n \text{ i.o.}) = 0$ .

To state the second part of the Borel-Cantelli lemma we need

**Definition 3.15.** Events  $A_1, A_2, \dots$ , in  $(\Omega, \mathcal{F}, P)$  will be called independent events if the random variables  $\chi_{A_1}, \chi_{A_2}, \dots$  are independent (see Problem 8).

II. The converse half. If  $A_n \in \mathcal{F}$  are independent events then  $\sum_1^\infty P(A_n) = \infty$  implies  $P(A_n \text{ i.o.}) = 1$ .

*Proof of I*

$$P(A_n \text{ i.o.}) = P\left(\lim_{m \rightarrow \infty} \bigcup_m A_n\right) = \lim_m P\left(\bigcup_m A_n\right) \leq \lim_m \left(\sum_m P(A_n)\right).$$

But obviously  $\sum_1^\infty P(A_n) < \infty$  implies that  $\sum_m P(A_n) \rightarrow 0$ , as  $m \rightarrow \infty$ .

*Proof of II*

$P\left(\bigcup_m A_n\right) = 1 - P\left(\bigcap_m A_n^c\right)$ . We show that  $P\left(\bigcap_m A_n^c\right) = 0$ . Because

the events  $\{A_n\}$  are independent,

$$P\left(\bigcap_m A_n^c\right) = \prod_m P(A_n^c) = \prod_m (1 - P(A_n)).$$

Use the inequality  $\log(1 - x) \leq -x$  to get

$$\log\left(\prod_m (1 - P(A_n))\right) = \sum_m \log(1 - P(A_n)) \leq -\sum_m P(A_n) = -\infty.$$

*Application 1.* In coin-tossing, let  $\mathbf{s}$  be any sequence  $k$ -long of  $H, T$ .

$$A_n = \{\omega; (\omega_n, \dots, \omega_{n+k-1}) = \mathbf{s}\}, \quad 0 < P(\text{Heads}) < 1.$$

**Proposition 3.16.**  $P(A_n \text{ i.o.}) = 1$ .

*Proof.* Let  $B_1 = \{\omega; (\omega_1, \dots, \omega_k) = \mathbf{s}\}$ ,  $B_2 = \{\omega; (\omega_{k+1}, \dots, \omega_{2k}) = \mathbf{s}\}, \dots$ . The difficulty is that the  $A_n$  are not independent events because of the overlap, for instance, between  $A_1$  and  $A_2$ , but the  $B_n$  are independent, and  $\{A_n \text{ i.o.}\} \supset \{B_n \text{ i.o.}\}$ . Now  $P(B_n) = P(B_1) > 0$ , so  $\sum_1^\infty P(B_n) = \infty$ , implying by 3.14(II) that  $P(B_n \text{ i.o.}) = 1$ .

Another way of putting this proposition is that in coin-tossing (biased or not), given any finite sequence of  $H, T$ 's, this sequence will occur an infinite number of times as the tossing continues, except on a set of sequences of probability zero.

*Application 2.* Again, in coin-tossing, let  $Y_i = \pm 1$ , as  $i$ th toss is  $H$  or  $T$ ,  $Z_n = Y_1 + \dots + Y_n$ . If  $Z_n = 0$ , we say that an equalization (or return to the origin) takes place at time  $n$ . Let  $A_n = \{Z_n = 0\}$ . Then  $\{A_n \text{ i.o.}\} = \{\omega; \text{an infinite number of equalizations occur}\}$ .

**Proposition 3.17.** If  $P(\text{Heads}) \neq \frac{1}{2}$ , then  $P(Z_n = 0 \text{ i.o.}) = 0$ .

*Proof.* Immediate, from the Borel-Cantelli lemma and the asymptotic expression for  $P(Z_n = 0)$ .

Another statement of 3.17 is that in biased coin-tossing, as we continue tossing, we eventually come to a last equalization and past this toss there are no more equalizations. What if the coin is fair?

**Theorem 3.18.** For a fair coin,  $P(Z_n = 0 \text{ i.o.}) = 1$ .

*Proof.* The difficulty, of course, is that the events  $A_n = \{Z_n = 0\}$  are not independent, so 3.14 is not directly applicable. In order to get around this, we manufacture a most pedestrian proof, which is typical of the way in which the Borel-Cantelli lemma is stretched out to cover cases of nonindependent events.

The idea of the proof is this; we want to apply the converse part of the Borel-Cantelli lemma, but in order to do this we can look only at the random variables  $X_k$  related to disjoint stretches of tosses. That is, if we

consider a subsequence  $n_1 < n_2 < n_3 < \dots$  of the integers, then any events  $\{C_k\}$  such that each  $C_k$  depends only on  $\{Y_{n_k+1}, Y_{n_k+2}, \dots, Y_{n_{k+1}}\}$  are independent events to which the Borel-Cantelli lemma applies. Suppose, for instance, that we select  $n_k < m_k < n_{k+1}$  and define

$$C_k = \{Y_{n_k+1} + \dots + Y_{m_k} \leq -n_k\} \cap \{Y_{m_k+1} + \dots + Y_{n_{k+1}} \geq m_k\}.$$

The purpose of defining  $C_k$  this way is that we know

$$Z_{n_k} = Y_1 + \dots + Y_{n_k} \leq n_k,$$

because each  $Y_i$  is  $\pm 1$ . Hence  $\omega \in C_k \Rightarrow Z_{m_k} \leq 0$ . Again  $Z_{m_k} \geq -m_k$ , so, in addition,

$$\{\omega \in C_k\} \Rightarrow Z_{n_{k+1}} \geq 0.$$

Therefore  $\{\omega \in C_k\} \Rightarrow \{Z_n = 0 \text{ at least once for } n_k + 1 \leq n \leq n_{k+1}\}$ . We have used here a standard trick in probability theory of considering stretches  $n_1, n_2, \dots$  so far apart that the effect of what happened previously to  $n_k$  is small as compared to the amount that  $Z_n$  can change between  $n_k$  and  $n_{k+1}$ . Now

$$\{C_k \text{ i.o.}\} \subset \{Z_n = 0 \text{ i.o.}\}.$$

So we need only to prove now that the  $n_k, m_k$  can be selected in such a way that  $\sum_1^\infty P(C_k) = \infty$ .

*Assertion:* Given any number  $\alpha, 0 < \alpha < 1$ , and integer  $k \geq 1, \exists$  an integer  $\varphi(k) \geq 1$  such that

$$P(|Z_{\varphi(k)}| < k) \leq \alpha.$$

*Proof.* We know that for any fixed  $j$ ,

$$P(Z_n = j) \rightarrow 0.$$

Hence for  $k$  fixed, as  $n \rightarrow \infty$ ,

$$\sum_{|j| < k} P(Z_n = j) \rightarrow 0.$$

Simply take  $\varphi(k)$  sufficiently large so that

$$\sum_{|j| < k} P(Z_{\varphi(k)} = j) \leq \alpha.$$

Define  $n_k, m_k$  as follows:  $n_1 = 1, m_k = n_k + \varphi(n_k), n_{k+1} = m_k + \varphi(m_k)$ . Compute  $P(C_k)$  as follows:

$$P(C_k) = P(Y_{n_k+1} + \dots + Y_{m_k} \leq -n_k)P(Y_{m_k+1} + \dots + Y_{n_{k+1}} \geq m_k).$$

By symmetry,

$$P(C_k) = \frac{1}{4}P(|Y_{n_k+1} + \dots + Y_{m_k}| \geq n_k)P(|Y_{m_k+1} + \dots + Y_{n_{k+1}}| \geq m_k).$$



Thus, since the distribution of the vector  $(Y_{i+1}, \dots, Y_{i+j})$  is the same as that of  $(Y_1, \dots, Y_j)$ ,

$$\begin{aligned} P(C_k) &= \frac{1}{4}P(|Y_1 + \dots + Y_{m_k - n_k}| \geq n_k)P(|Y_1 + \dots + Y_{n_{k+1} - m_k}| \geq m_k) \\ &= \frac{1}{4}P(|Y_1 + \dots + Y_{\varphi(n_k)}| \geq n_k)P(|Y_1 + \dots + Y_{\varphi(m_k)}| \geq m_k) \\ &\geq \frac{1}{4}(1 - \alpha)^2. \end{aligned} \quad \text{Q.E.D.}$$

This proof is a bit of a mess. Now let me suggest a much more exciting possibility. Suppose we can prove that  $P(Z_n = 0 \text{ at least once}) = 1$ . Now every time there is an equalization, everything starts all over again. That is, if  $Z_n = 0$ , then the game starts from the  $(n + 1)$  toss as though it were beginning at  $n = 0$ . Consequently, we are sure now to have at least one more equalization. Continue this argument now ad infinitum to conclude that  $P(Z_n = 0 \text{ at least once}) = 1 \Rightarrow P(Z_n = 0 \text{ i.o.}) = 1$ . We make this argument hold water when 3.18 is generalized in Section 7, and generalize it again in Chapter 7.

### Problems

6. Show, by using 3.14, that  $X_n \xrightarrow{P} X \Rightarrow \exists$  a subsequence  $X_{n_k}$  such that  $X_{n_k} \xrightarrow{\text{a.s.}} X$ .

7. Show, using 3.14, that if  $X_n - X_m \xrightarrow{P} 0$ ,  $\exists$  a random variable  $X$  such that  $X_n \xrightarrow{P} X$ . [Hint: Take  $\epsilon_k \downarrow 0$  and  $n_k$  such that for  $m, n \geq n_k$ ,

$$P(|X_n - X_m| > \epsilon_k) \leq \frac{1}{2^k}.$$

Now prove that there is a random variable  $X$  such that  $X_{n_k} \xrightarrow{\text{a.s.}} X$ .]

8. In order that events  $A_1, A_2, \dots$  be independent, show it is sufficient that

$$P(A_{i_1} \cap \dots \cap A_{i_m}) = P(A_{i_1}) \cdots P(A_{i_m})$$

for every finite subcollection  $A_{i_1}, \dots, A_{i_m}$ . [One interesting approach to the required proof is: Let  $\mathcal{D}$  be the smallest field containing  $A_1, \dots, A_N$ . Define  $Q$  on  $\mathcal{D}$  by  $Q(B_1 \cap \dots \cap B_N) = P(B_1) \cdots P(B_N)$ , where the sets  $B_k$  are equal to  $A_k$  or  $A_k^c$ . Use  $P(A_{i_1} \cap \dots \cap A_{i_m}) = P(A_{i_1}) \cdots P(A_{i_m})$  to show that  $P = Q$  on a class of sets to which 2.23 can be applied. Conclude that  $P = Q$  on  $\mathcal{D}$ .]

9. Use the strong law of large numbers in the form  $S_n/n \rightarrow p$  a.s. to prove 3.17.

10. Let  $X_1, X_2, \dots$  be independent identically distributed random variables. Prove that  $E|X_1| < \infty$  if and only if

$$P(|X_n| > n \text{ i.o.}) = 0.$$

(See Loève [108, p. 239].)

#### 4. THE RANDOM SIGNS PROBLEM

In Problem 5, it is shown for  $Y_1, Y_2, \dots$ , independent  $+1$  or  $-1$  with probability  $\frac{1}{2}$ , that the set  $\{\omega; \sum_1^n c_k Y_k \text{ converges}\}$  is a tail event. Therefore it has probability zero or one. The question now is to characterize the sequences  $\{c_n\}$  such that  $\sum_1^n c_k Y_k$  converges a.s.

This question is naturally arrived at when you look at the sequence  $1/n$ , that is,  $\sum 1/n$  diverges, but  $\sum (-1)^n 1/n$  converges. Now what happens if the signs are chosen at random? In general, look at the consecutive sums  $\sum_1^n X_k$  of any sequence  $X_1, X_2, \dots$  of independent random variables. The convergence set is again a tail event. When does it have probability one?

The basic result here is that in this situation convergence in probability implies the much stronger convergence almost surely.

**Theorem 3.19.** For  $X_1, X_2, \dots$  independent random variables,

$$(3.20) \quad \sum_1^n X_k \xrightarrow{P} \Leftrightarrow \sum_1^n X_k \xrightarrow{\text{a.s.}}$$

*Proof.* Proceeds by an important lemma which is due to Skorokhod, [125].

**Lemma 3.21.** Let  $S_1, \dots, S_N$  be successive sums of independent random variables such that  $\sup_{j \leq N} P(|S_N - S_j| > \alpha) = c < 1$ . Then

$$P\left(\sup_{j \leq N} |S_j| > 2\alpha\right) \leq \frac{1}{1-c} P(|S_N| > \alpha).$$

*Proof.* Let  $j^*(\omega) = \{\text{first } j \text{ such that } |S_j| > 2\alpha\}$ . Then

$$\begin{aligned} P\left(|S_N| > \alpha, \sup_{j \leq N} |S_j| > 2\alpha\right) &= \sum_{j=1}^N P(|S_N| > \alpha, j^* = j) \\ &\geq \sum_{j=1}^N P(|S_N - S_j| \leq \alpha, j^* = j). \end{aligned}$$

The set  $\{j^* = j\}$  is in  $\mathcal{F}(X_1, \dots, X_j)$ , and  $S_N - S_j$  is measurable  $\mathcal{F}(X_{j+1}, \dots, X_N)$ , so the last sum on the right above equals

$$\sum_{j=1}^N P(|S_N - S_j| \leq \alpha) P(j^* = j).$$

Use  $P(|S_N - S_j| \leq \alpha) \geq 1 - c$  to get

$$\begin{aligned} P\left(|S_N| > \alpha, \sup_{j \leq N} |S_j| > 2\alpha\right) &\geq (1-c) \sum_{j=1}^N P(j^* = j) \\ &= (1-c) P\left(\sup_{j \leq N} |S_j| > 2\alpha\right). \end{aligned}$$

The observation that

$$P(|S_N| > \alpha, \sup_{j \leq N} |S_j| > 2\alpha) \leq P(|S_N| > \alpha)$$

completes the proof.

To finish the proof of 3.19: If a sequence  $s_n$  of real numbers does not converge, then there exists an  $\epsilon > 0$  such that for every  $m$ ,

$$\sup_{n; n > m} |s_n - s_m| > \epsilon.$$

So if  $\sum_1^n X_k$  diverges with positive probability then there exists an  $\epsilon > 0$  and  $\delta > 0$  such that for every  $m$  fixed,

$$P\left(\sup_{n > m} \left| \sum_{k=m+1}^n X_k \right| > \epsilon\right) \geq \delta.$$

By (3.21), keeping  $m, N$  fixed

$$P\left(\sup_{m < n \leq N} \left| \sum_{k=m+1}^n X_k \right| > \epsilon\right) \leq \frac{1}{1 - C_{m,N}} P\left(\left| \sum_{k=m+1}^N X_k \right| > \frac{\epsilon}{2}\right),$$

where

$$C_{m,N} = \sup_{m < n \leq N} P\left(\left| \sum_{k=m+1}^n X_k \right| > \frac{\epsilon}{2}\right).$$

If  $\sum_1^n X_k$  is convergent in probability, then  $\sum_1^N X_k - \sum_1^m X_k \xrightarrow{P} 0$  as  $m, N \rightarrow \infty$ . Hence, as  $m, N \rightarrow \infty$ ,

$$P\left(\left| \sum_{k=m+1}^N X_k \right| > \frac{\epsilon}{2}\right) \rightarrow 0,$$

so we find that

$$C_{m,N} \rightarrow 0.$$

Taking first  $N \rightarrow \infty$ , conclude

$$\lim_m P\left(\sup_{n > m} \left| \sum_{k=m+1}^n X_k \right| > \epsilon\right) = 0.$$

This contradiction proves the theorem.

We can use convergence in second mean to get an immediate criterion.

**Corollary 3.22.** *If  $EX_k = 0$ , all  $k$ , and  $\sum_1^\infty EX_k^2 < \infty$ , then the sums  $\sum_1^n X_k$  converge a.s.*

In particular, for the random signs problem, mentioned at the beginning of this section, the following corollary holds.

**Corollary 3.23.** *A sufficient condition for the sums  $\sum_1^n c_k Y_k$  to converge a.s. is  $\sum_1^\infty c_k^2 < \infty$ .*

The open question is necessity. Marvelously enough, the converse of 3.23 is true, so that  $\sum c_k Y_k$  converges if and only if  $\sum c_k^2 < \infty$ . In fact, a partial converse of 3.22 holds.

**Theorem 3.24.** *Let  $X_1, X_2, \dots$  be independent random variables such that  $EX_k = 0$ , and  $|X_k| \leq \alpha < \infty$ , all  $k$ . Then  $\sum_1^n X_k$  converges a.s. implies  $\sum_1^\infty EX_k^2 < \infty$ .*

*Proof.* For any  $\lambda > 0$ , define  $n^*(\omega)$  by

$$n^*(\omega) = \begin{cases} \text{1st } n \text{ such that } \left| \sum_1^n X_k \right| > \lambda, \\ +\infty \text{ if no such } n \text{ exists,} \end{cases}$$

where  $n^*$  is an extended random variable. For any integers  $j < N$ , look at

$$\int_{\{n^*=j\}} \left( \sum_1^N X_k \right)^2 dP.$$

Since  $\{n^*=j\} \in \mathcal{F}(X_1, \dots, X_j)$ , then by independence, and  $EX_k = 0$ , all  $k$ ,

$$\int_{\{n^*=j\}} \left( \sum_1^N X_k \right)^2 dP = \int_{\{n^*=j\}} \left( \sum_1^j X_k \right)^2 dP + \sum_{j+1}^N \int_{\{n^*=j\}} X_k^2 dP.$$

On  $\{n^*=j\}$ ,  $\left| \sum_1^{j-1} X_k \right| \leq \lambda$ . Hence since  $|X_j| \leq \alpha$ ,  $\left| \sum_1^j X_k \right| \leq \lambda + \alpha$ .

And, by independence, for  $k > j$ ,

$$\int_{\{n^*=j\}} X_k^2 dP = P(n^*=j) EX_k^2.$$

Using these,

$$\int_{\{n^*=j\}} \left( \sum_1^N X_k \right)^2 dP \leq (\lambda + \alpha)^2 P(n^*=j) + P(n^*=j) \sum_1^N EX_k^2.$$

Sum on  $j$  from 1 up to  $N$  to get

$$\int_{\{n^* \leq N\}} \left( \sum_1^N X_k \right)^2 dP \leq (\lambda + \alpha)^2 P(n^* \leq N) + P(n^* \leq N) \cdot \sum_1^N EX_k^2.$$

Also,

$$\int_{\{n^* > N\}} \left( \sum_1^N X_k \right)^2 dP \leq \lambda^2 P(n^* > N) \leq (\lambda + \alpha)^2 P(n^* > N).$$

Adding this to the above inequality we get

$$\sum_1^N EX_k^2 \leq (\lambda + \alpha)^2 + P(n^* \leq N) \sum_1^N EX_k^2$$

or

$$P(n^* > N) \sum_1^N EX_k^2 \leq (\lambda + \alpha)^2.$$

Letting  $N \rightarrow \infty$ , we find that

$$P(n^* = \infty) \cdot \sum_1^\infty EX_k^2 \leq (\lambda + \alpha)^2.$$

But, since  $\sum_1^n X_k$  converges a.s., then there must exist a  $\lambda$  such that

$$P\left(\sup_n \left| \sum_1^n X_k \right| \leq \lambda\right) > 0,$$

implying  $P(n^* = \infty) > 0$  and  $\sum_1^\infty EX_k^2 < \infty$ .

The results of 3.24 can be considerably sharpened. But why bother; elegant necessary and sufficient conditions exist for the convergence of sums  $\sum_1^n X_k$  where the only assumption made is that the  $X_k$  are independent. This is the "three-series" theorem of Kolmogorov (see Loève [108, p. 237]). More on this will appear in Chapter 9. Kac [82] has interesting analytic proofs of 3.23 and its converse.

### Problems

11. Let  $X_1, X_2, \dots$  be independent, and  $X_k \geq 0$ . If for some  $\delta, 0 < \delta < 1$ , there exists an  $x$  such that

$$\int_{\{X_k > x\}} X_k dP \leq \delta EX_k,$$

for all  $k$ , show that

$$\sum_1^\infty X_k < \infty \text{ a.s.} \Rightarrow \sum_1^\infty EX_k < \infty.$$

Give an example to show that in general,  $X_1, X_2, \dots$  independent non-negative random variables and  $\sum_1^\infty X_k < \infty$  a.s. does not imply that

$$\sum_1^\infty EX_k < \infty.$$

**12.** Let  $Y_1, Y_2, \dots$  be a process. We will say that the integer-valued random variables  $m_1, m_2, \dots$  are *optional skipping* variables if

- i)  $1 < m_1 < m_2 < \dots$ ,
- ii)  $\{m_k = j\} \in \mathcal{F}(Y_1, \dots, Y_{j-1})$

(i.e., the decision as to which game to play next depends only on the previous outcomes). Denote  $\check{Y}_k = Y_{m_k}$ . Show that

- a) If the  $Y_1, Y_2, \dots$  are independent and identically distributed then the sequence  $\check{Y}_1, \check{Y}_2, \dots$  has the same distribution as  $Y_1, Y_2, \dots$ .
- b) For  $Y_1, Y_2, \dots$  as in (a), show that the sequence  $Y_{m_1}, Y_{m_1+1}, Y_{m_1+2}, \dots$  has the same distribution as  $Y_1, Y_2, \dots$ .
- c) Give an example where the  $Y_1, Y_2, \dots$  are independent, but the  $\check{Y}_1, \check{Y}_2, \dots$  are not independent.

## 5. THE LAW OF PURE TYPES

Suppose that  $X_1, X_2, \dots$  are independent and  $\sum_1^n X_k$  converges a.s. What can be said about the distribution of the limit  $X = \sum_1^\infty X_k$ ? In general, very little! In the nature of things the distribution of  $X$  can be anything. For example, let  $X_k = 0, k > 1$ , then  $X = X_1$ . There is one result available here which is an application of the Kolmogorov zero-one law and remarkable for its simplicity and elegance.

**Definition 3.25.** A random variable  $X$  is said to have a distribution of pure type if either

- i) There is a countable set  $D$  such that  $P(X \in D) = 1$ ,
- ii)  $P(X = x) = 0$  for every  $x \in R^{(1)}$ , but there is a set  $D \in \mathcal{B}_1$  of Lebesgue measure zero such that  $P(X \in D) = 1$ , or
- iii)  $P(X \in dx) \ll l(dx)$  (Lebesgue measure)

[Recall that  $\mu \ll \nu$  for two measures  $\mu, \nu$  denotes  $\mu$  absolutely continuous with respect to  $\nu$ ; see Appendix A.29].

**Theorem 3.26** (Jessen-Wintner law of pure types [78]). Let  $X_1, X_2, \dots$  be independent random variables such that

- i)  $\sum_1^n X_k \rightarrow X$  a.s.,
- ii) For each  $k$ , there is a countable set  $F_k$  such that  $P(X_k \in F_k) = 1$ .

Then the distribution of  $X$  is of pure type.

*Proof.* Let  $F = \bigcup_{k \geq 1} F_k$ . Take  $G$  to be the smallest additive group in  $R^{(1)}$  containing  $F$ .  $G$  consists of all numbers of the form  $m_1 x_1 + \dots + m_j x_j$ ,

$x_1, \dots, x_j \in F$  and  $m_1, \dots, m_j$  integers.  $F$  is countable, hence  $G$  is countable. For any set  $B \subset R^{(1)}$  write

$$G \oplus B = \{x \in R^{(1)}; x = x_1 + x_2, x_1 \in G, x_2 \in B\}.$$

Note that

- i)  $B$  countable  $\Rightarrow G \oplus B$  countable,
- ii)  $B \in \mathfrak{B}_1 \Rightarrow G \oplus B \in \mathfrak{B}_1$ ,
- iii)  $B \in \mathfrak{B}_1, l(B) = 0 \Rightarrow l(G \oplus B) = 0$ .

For  $B \in \mathfrak{B}_1$ , and  $C = \{\omega, \sum_1^n X_k \text{ converges}\}$ , consider the event

$$A = \{X \in G \oplus B\} \cap C.$$

The point is that  $A$  is a tail event. Because if  $x_1 - x_2 \in G$ , then

$$x_1 \in G \oplus B \Leftrightarrow x_2 \in G \oplus B.$$

But  $X - \sum_n X_k \in G$  for all  $\omega$  in  $C$ . Hence

$$A = \left\{ \sum_n X_k \in G \oplus B \right\} \cap C, n = 1, 2, \dots$$

By the zero-one law  $P(A) = 0$  or  $1$ . This gives the alternatives:

- a) Either there is a countable set  $D$  such that  $P(X \in D) = 1$  or  $P(X \in G \oplus B) = 0$ , hence  $P(X \in B) = 0$ , for all countable sets  $B$ .
- b) If the latter in (a) holds, then either there is a set  $D \in \mathfrak{B}_1$  such that  $l(D) = 0$  and  $P(X \in D) = 1$  or  $P(X \in G \oplus B) = 0$ , for all  $B \in \mathfrak{B}_1$  such that  $l(B) = 0$ .
- c) In this latter case  $B \in \mathfrak{B}_1, l(B) = 0 \Rightarrow P(X \in B) = 0$ , that is, the distribution is absolutely continuous with respect to  $l(dx)$ .

Theorem 3.26 gives no help as to which type the distribution of the limit random variable belongs to. In particular, for  $Y_1, \dots$  independent  $\pm 1$  with probability  $\frac{1}{2}$ , the question of the type of the distribution of the sums  $\sum_1^n c_n Y_n$  is open. Some important special cases are given in the following problems.

### Problems

13. Show that  $P\left(\sum_1^\infty Y_k/2^k \in dx\right)$  is Lebesgue measure on  $[0, 1]$ . [Recall the analytic model for coin-tossing.]

14. If  $X$  and  $Y$  are independent random variables, use Problem 2 to show that  $P(X \in dx) \ll l(dx) \Rightarrow P(X + Y \in dx) \ll l(dx)$ .

15. Use Problems 13 and 14 to show that the distribution of

$$\sum_1^{\infty} Y_k/k \text{ is } \ll l(dx).$$

16. Show that if independent random variables  $X_1, X_2, \dots$  take values in a countable set  $F$ , and if there are constants  $\alpha_n \in F$  such that

$$\sum_1^{\infty} P(X_n \neq \alpha_n) < \infty \quad \text{and} \quad \sum_1^{\infty} \alpha_n < \infty,$$

then the sum  $X = \sum_1^{\infty} X_n$  has distribution concentrated on a countable number of points. (The converse is also true; see Lévy [101].)

## 6. THE LAW OF LARGE NUMBERS FOR INDEPENDENT RANDOM VARIABLES

From the random signs problem, by some juggling, we can generalize the law of large numbers for independent random variables.

**Theorem 3.27.** Let  $X_1, X_2, \dots$  be independent random variables,  $EX_k = 0$ ,  $EX_k^2 < \infty$ . Let  $b_n \geq 0$  converge up to  $+\infty$ . If  $\sum_1^{\infty} EX_k^2/b_k^2 < \infty$ , then

$$\frac{X_1 + \dots + X_n}{b_n} \xrightarrow{\text{a.s.}} 0.$$

*Proof.* To prove this we need:

**Kronecker's Lemma 3.28.** Let  $x_1, x_2, \dots$  be a sequence of real numbers such that  $\sum_1^n x_k \rightarrow s$  finite. Take  $b_n \uparrow \infty$ , then

$$\frac{1}{b_n} \sum_1^n b_k x_k \rightarrow 0.$$

*Proof.* Let  $r_n = \sum_{n+1}^{\infty} x_k$ ,  $r_0 = s$ ; then  $x_n = r_{n-1} - r_n$ ,  $n = 1, 2, \dots$ , and

$$\begin{aligned} \sum_1^n b_k x_k &= \sum_1^n b_k (r_{k-1} - r_k) = \sum_0^{n-1} b_{k+1} r_k - \sum_1^n b_k r_k \\ &= \sum_1^{n-1} (b_{k+1} - b_k) r_k + b_1 s - b_n r_n \end{aligned}$$

Thus

$$(3.29) \quad \left| \sum_1^n b_k x_k \right| \leq \sum_1^{n-1} (b_{k+1} - b_k) |r_k| + b_1 |s| + b_n |r_n|.$$



For any  $\epsilon > 0$ , take  $N$  such that  $|r_k| \leq \epsilon$  for  $k \geq N$ . Then letting  $\bar{r} = \max_{n \geq 1} |r_n|$  we get

$$\begin{aligned} \sum_1^{n-1} (b_{k+1} - b_k) |r_k| &\leq \sum_{k=1}^{N-1} (b_{k+1} - b_k) |r_k| + \epsilon \sum_N^{n-1} (b_{k+1} - b_k) \\ &\leq \bar{r}(b_N - b_1) + \epsilon(b_n - b_N). \end{aligned}$$

Divide (3.29) by  $b_n$ , and take lim sup, noting that  $b_N/b_n \rightarrow 0$ ,  $|r_n| \rightarrow 0$ , to get

$$\overline{\lim}_n \left| \frac{1}{b_n} \sum_1^n b_k x_k \right| \leq \epsilon.$$

To prove 3.27, by Kronecker's lemma, if  $\sum_1^n (X_k/b_k)$  converges a.s., then

$$\frac{1}{b_n} \sum X_k \rightarrow 0 \text{ a.s.}$$

By 3.22, it is sufficient that

$$\sum_1^\infty E(X_k/b_k)^2 = \sum_1^\infty EX_k^2/b_k^2 < \infty.$$

Q.E.D.

As a consequence of 3.27, if the  $X_1, X_2, \dots$  are identically distributed,  $EX_1 = 0$ , and  $EX_1^2 = \sigma^2 < \infty$ , then

$$\sum_1^\infty \frac{1}{b_k^2} < \infty \Rightarrow \frac{X_1 + \dots + X_n}{b_n} \rightarrow 0 \text{ a.s.}$$

This is stronger than 1.21, which gives the same conclusion for  $b_n = n$ . For example, we could take  $b_n = n^{1/2+\epsilon}$ , any  $\epsilon > 0$ ; or  $b_n = \sqrt{n \log n}$ . But the strong law of large numbers is basically a first-moment theorem.

**Theorem 3.30.** Let  $X_1, X_2, \dots$  be independent and identically distributed random variables; if  $E|X_1| < \infty$  then

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{a.s.}} EX_1;$$

if  $E|X_1| = \infty$ , then the above averages diverge almost everywhere.

*Proof.* In order to apply 3.27 define truncated random variables  $\tilde{X}_n$  by

$$\tilde{X}_n = \begin{cases} X_n, & \text{if } |X_n| \leq n, \\ 0, & |X_n| > n. \end{cases}$$

By Problem 10,  $P(|X_n| > n \text{ i.o.}) = 0$ . Hence (3.30) is equivalent to

$$\frac{\tilde{X}_1 + \dots + \tilde{X}_n}{n} \xrightarrow{\text{a.s.}} EX_1.$$

But  $E\tilde{X}_n \rightarrow EX_1$ , so (3.30) will follow if 3.27 can be applied to show that

$$\frac{1}{n} \sum_{k=1}^n (\tilde{X}_k - E\tilde{X}_k) \rightarrow 0.$$

Since  $E(\tilde{X}_k - E\tilde{X}_k)^2 \leq E\tilde{X}_k^2$ , it is sufficient to show that

$$\sum_{k=1}^{\infty} \frac{1}{n^2} E\tilde{X}_k^2 = \sum_{k=1}^{\infty} \frac{1}{n^2} \int_{|x| \leq n} x^2 F(dx) < \infty.$$

This follows from writing the right-hand side as

$$\sum_{n=1}^{\infty} \sum_{k=1}^n \frac{1}{n^2} \int_{k-1 < |x| \leq k} x^2 F(dx).$$

Interchange order of summation, and use  $\sum_{k=1}^{\infty} 1/n^2 < 2/k$ ,  $k \geq 1$ , to get

$$\sum_{k=1}^{\infty} \frac{1}{n^2} E\tilde{X}_k^2 \leq \sum_{k=1}^{\infty} \frac{2}{k} \int_{k-1 < |x| \leq k} x^2 F(dx) \leq 2E|X_1|.$$

For the converse, suppose that  $S_n/n$  converges on a set of positive probability. Then it converges a.s. The contradiction is that

$$\frac{X_n}{n} = \frac{S_n}{n} - \left( \frac{n-1}{n} \right) \frac{S_{n-1}}{n-1}$$

must converge a.s. to zero, implying  $P(|X_n| > n \text{ i.o.}) = 0$ . This is impossible by Problem 10.

## 7. RECURRENCE OF SUMS

Through this section let  $X_1, X_2, \dots$  be a sequence of independent, identically distributed random variables. Form the successive sums  $S_1, S_2, \dots$

**Definition 3.31.** For  $x \in R^{(1)}$ , call  $x$  a recurrent state if for every neighborhood  $I$  of  $x$ ,  $P(S_n \in I \text{ i.o.}) = 1$ .

The problem is to characterize the set of recurrent states. In coin-tossing, with  $X_1, X_2, \dots$  equaling  $\pm 1$  with probability  $p, q$ , if  $p \neq \frac{1}{2}$ , then  $P(S_n = 0 \text{ i.o.}) = 0$ . In fact, the strong law of large numbers implies that for any state  $j$ ,  $P(S_n = j \text{ i.o.}) = 0$ —no states are recurrent. For fair coin-tossing, every time  $S_n$  returns to zero, the probability of entering the state  $j$  is the same as it was at the start when  $n = 0$ . It is natural to surmise that in this case  $P(S_n = j \text{ i.o.}) = 1$  for all  $j$ . But we can use this kind of reasoning for any distribution, that is, if there is any recurrent state, then all states should be recurrent.

**Definition 3.32.** Say that a random variable  $X$  is distributed on the lattice  $L_d = \{nd\}$ ,  $n = 0, \pm 1, \dots$ ,  $d$  any real number  $> 0$ , if  $\sum_n P(X = nd) = 1$  and there is no smaller lattice having this property. If  $X$  is not distributed on any lattice, it is called nonlattice. In this case, say that it is distributed on  $L_0$ , where  $L_0 = R^{(1)}$ .

**Theorem 3.33.** If  $X_1, X_2, \dots$  are distributed on  $L_d$ ,  $d \geq 0$ , then either every state in  $L_d$  is recurrent, or no states are recurrent.

*Proof.* Let  $G$  be the set of recurrent points. Then  $G$  is closed. Because  $x_n \in G$ ,  $x_n \rightarrow x$  implies that for every neighborhood  $I$  of  $x$ ,  $x_n \in I$  for  $n$  sufficiently large. Hence  $P(S_n \in I \text{ i.o.}) = 1$ .

Define  $y$  to be a possible state if for every neighborhood  $I$  of  $y$ ,  $\exists k$  such that  $P(S_k \in I) > 0$ . I assert that

$$x \text{ recurrent, } y \text{ possible} \Rightarrow x - y \text{ recurrent.}$$

To show this, take any  $\epsilon > 0$ , and  $k$  such that  $P(|S_k - y| < \epsilon) > 0$ . Then  $P(|S_n - x| < \epsilon \text{ finitely often})$

$$\begin{aligned} &\geq P(|S_k - y| < \epsilon, |S_{k+n} - S_k - (x - y)| < 2\epsilon \text{ finitely often}) \\ &= P(|S_k - y| < \epsilon)P(|S_n - (x - y)| < 2\epsilon \text{ finitely often}). \end{aligned}$$

The left side is zero, implying

$$P(|S_n - (x - y)| < 2\epsilon \text{ finitely often}) = 0.$$

If  $G$  is not empty, it contains at least one state  $x$ . Since every recurrent state is a possible state,  $x - x = 0 \in G$ . Hence  $G$  is a group, and therefore is a closed subgroup of  $R^{(1)}$ . But the only closed subgroups of  $R^{(1)}$  are the lattices  $L_d$ ,  $d \geq 0$ . For every possible state  $y$ ,  $0 - y \in G \Rightarrow y \in G$ . For  $d > 0$ , this implies  $L_d \subset G$ , hence  $L_d = G$ . If  $X_1$  is non-lattice,  $G$  cannot be a lattice, so  $G = R^{(1)}$ .

A criterion for which alternative holds is established in the following theorem.

**Theorem 3.34.** Let  $X_1, X_2, \dots$  be distributed on  $L_d$ ,  $d \geq 0$ . If there is a finite interval  $J$ ,  $J \cap L_d \neq \emptyset$  such that  $\sum_1^\infty P(S_n \in J) < \infty$ , then no states are recurrent. If there is a finite interval  $J$  such that  $\sum_1^\infty P(S_n = J) = \infty$ , then all states in  $L_d$  are recurrent.

*Proof.* If  $\sum_1^\infty P(S_n \in J) < \infty$ , use the Borel-Cantelli lemma to get  $P(S_n \in J \text{ i.o.}) = 0$ . There is at least one state in  $L_d$  that is not recurrent, hence none are. To go the other way, we come up against the same difficulty as in proving 3.18, the successive sums  $S_1, S_2, \dots$  are not independent. Now we make essential use of the idea that every time one of the sums

$S_1, S_2, \dots$  comes back to the neighborhood of a state  $x$ , the subsequent process behaves nearly as if we had started off at the state  $x$  at  $n = 0$ . If  $\sum_1^\infty P(S_n \in J) = \infty$ , for any  $\epsilon > 0$  and less than half the length of  $J$ , there is a subinterval  $I = (x - \epsilon, x + \epsilon) \subset J$  such that  $\sum_1^\infty P(S_n \in I) = \infty$ . Define sets

$$A_k = \begin{cases} \{S_k \in I, S_{n+k} \notin I, n = 1, 2, \dots\}, & k \geq 1, \\ \{S_n \notin I, n = 1, 2, \dots\}, & k = 0. \end{cases}$$

$A_k$  is the set on which the last visit to  $I$  occurred at the  $k$ th trial. Then

$$\{S_n \in I \text{ finitely often}\} = \bigcup_0^\infty A_k.$$

The  $A_k$  are disjoint, hence

$$P(S_n \in I \text{ finitely often}) = \sum_0^\infty P(A_k).$$

For  $k \geq 1$ ,

$$P(A_k) \geq P(S_k \in I, |S_{n+k} - S_k| \geq 2\epsilon, n = 1, 2, \dots).$$

Use independence, then identical distribution of the  $\{X_n\}$  to get

$$\begin{aligned} P(A_k) &\geq P(S_k \in I)P(|S_{n+k} - S_k| \geq 2\epsilon, n = 1, 2, \dots) \\ &= P(S_k \in I)P(|S_n| \geq 2\epsilon, n = 1, 2, \dots). \end{aligned}$$

This inequality holds for all  $k \geq 1$ . Thus

$$P(S_n \in I \text{ finitely often}) \geq P(|S_n| \geq 2\epsilon, n = 1, 2, \dots) \sum_1^\infty P(S_k \in I).$$

Since  $\sum_1^\infty P(S_k \in I) = \infty$ , we conclude that for every  $\epsilon > 0$

$$(3.35) \quad P(|S_n| \geq 2\epsilon, n = 1, 2, \dots) = 0.$$

Now take  $I = (-\epsilon, +\epsilon)$ , and define the sets  $A_k$  as above. Denote  $I_\delta = (-\delta, +\delta)$ , so that

$$A_k = \lim_{\delta \uparrow \epsilon} \{S_k \in I_\delta, S_{n+k} \notin I, n = 1, 2, \dots\}, \quad k \geq 1.$$

Since the sequence of sets is monotone,

$$P(A_k) = \lim_{\delta \uparrow \epsilon} P(S_k \in I_\delta, S_{n+k} \notin I, n = 1, 2, \dots), \quad k \geq 1.$$

Now use (3.35) in getting

$$\begin{aligned} &P(S_k \in I_\delta, S_{n+k} \notin I, n = 1, 2, \dots) \\ &\leq P(S_k \in I_\delta, |S_{n+k} - S_k| \geq \epsilon - \delta, n = 1, \dots) \\ &= P(S_k \in I_\delta)P(|S_n| \geq \epsilon - \delta, n = 1, \dots) = 0 \end{aligned}$$

to conclude  $P(A_k) = 0, k \geq 1$ . Use (3.35) directly to establish

$$P(A_0) = P(|S_n| \geq \epsilon, n = 1, 2, \dots) = 0.$$

Therefore  $P(S_n \in I \text{ finitely often}) = 0$ , and the sums  $S_n$  enter every neighborhood of the origin infinitely often with probability one. So the origin is a recurrent state, and consequently all states in  $L_a$  are recurrent. Q.E.D.

Look again at the statement of this theorem. An immediate application of the Borel-Cantelli lemma gives a zero-one property:

**Corollary 3.36.** *Either*

$$P(S_n \in I \text{ i.o.}) = 1$$

*for all finite intervals  $I$  such that  $L_a \cap I \neq \emptyset$ ; or*

$$P(S_n \in I \text{ i.o.}) = 0$$

*for all such  $I$ .*

**Definition 3.37.** *If the first alternative in 3.36 holds, call the process  $S_1, S_2, \dots$  recurrent. If the second holds, call it transient.*

A quick corollary of 3.34 is a proof that fair coin-tossing is recurrent. Simply use the estimate  $P(S_{2n} = 0) \sim 1/\sqrt{\pi n}$  to deduce that

$$\sum_1^\infty P(S_n = 0)$$

diverges.

The criterion for recurrence given in 3.34 is difficult to check directly in terms of the distribution of  $X_1, X_2, \dots$ . A slightly more workable expression will be developed in Chapter 8. There is one important general result, however. If

$$E|X_1| < \infty \quad \text{and} \quad EX_1 \neq 0,$$

then by the law of large numbers, the sums are transient. If  $EX_1 = 0$ , the issue is in doubt. All that is known is that  $S_n = o(n)$  [ $o(n)$  denoting small order of  $n$ ]. There is no reason why the sums should behave as regularly as the sums in coin-tossing. (See Problem 17 for a particularly badly-behaved example of successive sums with zero means.) But, at any rate,

**Theorem 3.38.** *If  $EX_1 = 0$ , then the sums  $S_1, S_2, \dots$  are recurrent.*

*Proof.* First, we need to prove

**Proposition 3.39.** *If  $I$  is any interval of length  $a$ , then*

$$\sum_1^\infty P(S_n \in I) \leq 1 + \sum_1^\infty P(|S_n| \leq a).$$

*Proof.* Denote  $N = \sum_1^\infty \chi_I(S_n)$ , so that  $N$  counts the number of times that the sums enter the interval  $I$ . Define an extended random variable  $n^*$  by

$$n^* = \begin{cases} \text{1st } n \text{ such that } S_n \in I, \\ \infty \text{ if } S_n \notin I, n = 1, 2, \dots \end{cases}$$

$$EN = \sum_1^\infty \int_{\{n^*=k\}} N dP.$$

On  $\{n^* = k\}$ ,  $\chi_I(S_n) = 0$ ,  $n < k$ , and  $\chi_I(S_k) = 1$ . Thus, on  $\{n^* = k\}$ , denoting by  $I - y$  the interval  $I$  shifted left a distance  $y$ ,

$$\begin{aligned} N &\leq 1 + \sum_{n=1}^\infty \chi_{I-S_k}(S_{n+k} - S_k) \\ &\leq 1 + \sum_{n=1}^\infty \chi_{[-a, +a]}(S_{n+k} - S_k). \end{aligned}$$

Since  $\{n^* = k\} \in \mathcal{F}_1(X_1, \dots, X_k)$ ,

$$EN \leq P(n^* < \infty) \left[ 1 + \sum_1^\infty P(|S_n| \leq a) \right].$$

We use 3.39 to prove 3.38 as follows: For any positive integer  $M$ ,

$$\begin{aligned} \sum_1^\infty P(-M < S_n \leq M) &= \sum_{k=-M}^{M-1} \sum_1^\infty P(S_n \in (k, k+1)) \\ &\leq 2M \left( 1 + \sum_1^\infty P(|S_n| \leq 1) \right). \end{aligned}$$

Hence

$$(3.40) \quad \overline{\lim} \frac{1}{2M} \sum_1^\infty P(|S_n| < M) \leq 1 + \sum_1^\infty P(|S_n| \leq 1).$$

The strong law of large numbers implies the weaker result  $S_n/n \xrightarrow{P} 0$ , or  $P(|S_n| < \epsilon n) \rightarrow 1$  for every  $\epsilon > 0$ . Fix  $\epsilon$ , and take  $m$  so that  $P(|S_n| < \epsilon n) > \frac{1}{2}$ ,  $n > m$ . Then  $P(|S_n| < M) > \frac{1}{2}$ ,  $m \leq n < M/\epsilon$ , which gives

$$\begin{aligned} \sum_1^\infty P(|S_n| < M) &\geq \frac{1}{2}(M/\epsilon - m), \\ (1/2M) \sum_1^\infty P(|S_n| < M) &\geq 1/4\epsilon - m/4M. \end{aligned}$$

Substituting this into (3.40), we get

$$1 + \sum_1^\infty P(|S_n| \leq 1) \geq 1/4\epsilon.$$

Since  $\epsilon$  is arbitrary, conclude that

$$\sum_1^{\infty} P(|S_n| \leq 1) = \infty.$$

By 3.34 the sums are recurrent. Q.E.D.

### Problems

**17.** Unfavorable “fair” game, due to Feller [59, Vol. I, p. 246]. Let  $X_1, X_2, \dots$  be independent and identically distributed, and take values in  $\{0, 2, 2^2, 2^3, \dots\}$  so that

$$P(X_1 = 2^k) = \frac{1}{2^k k(k+1)},$$

and define  $P(X_1 = 0)$  to make the sum unity. Now  $EX_1 = 1$ , but show that for every  $\epsilon > 0$ ,

$$P\left(S_n - n < -\frac{(1 - \epsilon)n}{\log_2 n}\right) \rightarrow 1.$$

**18.** Consider  $k$  fair coin-tossing games being carried on independently of each other, giving rise to the sequence of random variables

$$\begin{aligned} &Y_1^{(1)}, Y_2^{(1)}, \dots, \\ &Y_1^{(2)}, Y_2^{(2)}, \dots, \\ &\vdots \\ &Y_1^{(k)}, Y_2^{(k)}, \dots, \end{aligned}$$

where  $Y_n^{(j)}$  is  $\pm 1$  as the  $n$ th outcome of the  $j$ th game is  $H$  or  $T$ . Let  $Z_n^{(j)} = Y_1^{(j)} + \dots + Y_n^{(j)}$ , and plot the progress of each game on one axis of  $R^{(k)}$ . The point described is  $Z_n = (Z_n^{(1)}, \dots, Z_n^{(k)}) = Y_1 + \dots + Y_n$ , where  $Y_1$  takes any one of the values  $(\pm 1, \dots, \pm 1)$  with probability  $1/2^k$ . Denote  $\mathbf{0} = (0, 0, \dots, 0)$ . Now  $Z_n = \mathbf{0}$  only if equalization takes place in all  $k$  games simultaneously. Show that

$$P(Z_n = \mathbf{0} \text{ i.o.}) = \begin{cases} 1, & k = 1, 2, \\ 0, & k \geq 3. \end{cases}$$

### 8. STOPPING TIMES AND EQUIDISTRIBUTION OF SUMS

Among the many nice applications of the law of large numbers, I am going to pick one. Suppose that  $X_1, X_2, \dots$  are distributed on the lattice  $L_d$  and the sums are recurrent. For any interval  $I$  such that  $I \cap L_d \neq \emptyset$ , the number of  $S_1, \dots, S_n$  falling into  $I$  goes to  $\infty$ . Denote this number by  $N_n(I)$ . Then

$N_n/n$ , the average number of landings in  $I$  per unit time, goes to zero in general (see Problem 20).

An interesting result is that the points  $S_1, \dots, S_n$  become a.s. uniformly distributed in the sense that for any two finite intervals

$$\frac{N_n(I_1)}{N_n(I_2)} \xrightarrow{\text{a.s.}} \frac{\|I_1\|}{\|I_2\|}.$$

(In the lattice case, define  $\|I\|$  as the number of points of  $L_d$  in  $I \cap L_d$ ). This equidistribution is clearly a strong property. The general proof is not elementary (see Harris and Robbins [69]). But there is an interesting proof in the lattice case which introduces some useful concepts. The idea is to look at the number of landings of the  $S_n$  sequence in  $I$  between successive zeros of  $S_n$ .

**Definition 3.41.** A positive, integer-valued random variable  $n^*$  is called a stopping time for the sums  $S_1, S_2, \dots$  if

$$\{n^* = k\} \in \mathcal{F}(S_1, \dots, S_k).$$

The field of events  $\mathcal{F}(S_k, k \leq n^*)$  depending on  $S_n$  up to time of stopping consists of all  $A \in \mathcal{F}(X)$  such that

$$A \cap \{n^* = k\} \in \mathcal{F}(S_1, \dots, S_k).$$

For example, in recurrent case,  $d = 1$ , let  $n^*$  be the first entry of the sums  $\{S_n\}$  into the state  $j$ . This is a stopping time. More important, once at state  $j$ , the process continues by adding independent random variables, so that  $S_{n^*+k} - S_{n^*}$  should have the same distribution as  $S_k$  and be independent of anything that happened up to time  $n^*$ .

**Proposition 3.42.** If  $n^*$  is a stopping time, then the process  $\check{S}_k = S_{n^*+k} - S_{n^*}$ ,  $k = 1, \dots$  has the same distribution as  $S_k$ ,  $k = 1, \dots$  and  $\mathcal{F}(\check{S}_1, \check{S}_2, \dots)$  is independent of  $\mathcal{F}(S_k, k \leq n^*)$ .

*Proof.* Let  $A \in \mathcal{F}(S_k, k \leq n^*)$ ,  $B \in \mathcal{B}_\infty$ , and write

$$P(\check{S} \in B, A) = \sum_{n=1}^{\infty} P(\check{S} \in B, A, n^* = n).$$

On the set  $\{n^* = n\}$ , the  $\check{S}_k$  process is equal to the  $S_{k+n} - S_n$  process, and  $A \cap \{n^* = n\} \in \mathcal{F}(S_1, \dots, S_n)$ . Since  $S_{k+n} - S_n$ ,  $k = 1, 2, \dots$  has the same distribution as  $S_k$ ,  $k = 1, 2, \dots$

$$P(\check{S} \in B, A) = P(S \in B) \sum_{n=1}^{\infty} P(A, n^* = n) = P(S \in B)P(A).$$

Note that  $n^*$  itself is measurable  $\mathcal{F}(S_k, k \leq n^*)$ .



**Definition 3.43.** The times of the zeros of  $S_n$  are defined by

$$\begin{aligned} R_1 &= \min \{n; S_n = 0, n > 0\}, \\ R_2 &= \min \{n; S_n = 0, n > R_1\}, \\ &\vdots \\ R_k &= \min \{n; S_n = 0, n > R_{k-1}\}. \end{aligned}$$

$R_k$  is called the  $k$ th occurrence time of  $\{S_n = 0\}$ . The times between zeros are defined by  $T_1 = R_1, T_2 = R_2 - R_1, \dots$

The usefulness of these random variables is partially accounted for by

**Proposition 3.44.** If  $P(S_n = 0 \text{ i.o.}) = 1$ , then the  $T_1, T_2, \dots$  are independent and identically distributed random variables.

*Proof.*  $T_1$  is certainly a stopping time. By 3.42,  $\check{S}_k = S_{k+T_1} - S_{T_1}$  has the same distribution as  $S_k$ , but this process is independent of  $T_1$ . Thus,  $T_2$ , which is the first equalization time for the  $\check{S}_k$  process, is independent of  $T_1$  and has the same distribution. Repeat this argument for  $k = 3, \dots$

**Theorem 3.45.** Let  $X_1, X_2, \dots$  have lattice distance one, and  $P(S_n = 0 \text{ i.o.}) = 1$ . Then, for any two states,  $j, l$ ,

$$\frac{N_n(j)}{N_n(l)} \xrightarrow{\text{a.s.}} 1.$$

*Proof.* Let  $R_1, R_2, \dots$  be the times of successive returns to the origin,  $T_1, T_2, \dots$  the times between return. The  $T_1, T_2, \dots$  are independent and identically distributed. Let  $M_1(j)$  be the number of landings in  $j$  before the first return to the origin,  $M_2(j)$ , the number between the first and second returns, etc. The  $M_1, M_2, \dots$  are similarly independent and identically distributed (see Problem 22). The law of large numbers could be applied to  $(M_1 + \dots + M_k)/k$  if we knew something about  $EM_1(j)$ . Denote  $\pi(j) = EM_1(j)$ , and assume for the moment that for all  $j \in L_1$ ,  $0 < \pi(j) < \infty$ . Since  $M_1(j) + \dots + M_k(j) = N_{R_k}(j)$ ,

$$\frac{N_{R_k}(j)}{N_{R_k}(l)} \xrightarrow{\text{a.s.}} \frac{\pi(j)}{\pi(l)}, \quad j \neq 0, \quad l \neq 0.$$

This gives convergence of  $N_n(j)/N_n(l)$  along a random subsequence. To get convergence over the full sequence, write

$$\begin{aligned} \max_{0 \leq m \leq T_{k+1}} \left| \frac{N_{R_k+m}(j)}{N_{R_k+m}(l)} - \frac{N_{R_k}(j)}{N_{R_k}(l)} \right| &\leq \frac{N_{R_{k+1}}(j) - N_{R_k}(j)}{N_{R_k}(l)} \\ &\quad + \frac{N_{R_k}(j)[N_{R_{k+1}}(l) - N_{R_k}(l)]}{[N_{R_k}(l)]^2}. \end{aligned}$$

Dividing the top and bottom of the first term on the right by  $k$  we have that  $\overline{\lim}_k$  of that term is given by

$$\frac{1}{EM_1(l)} \overline{\lim}_k \left( \frac{M_{k+1}(j)}{k} \right).$$

By Problem 10, this term is a.s. zero. The second term is treated similarly to get

$$(3.46) \quad \frac{N_n(j)}{N_n(l)} \xrightarrow{\text{a.s.}} \frac{\pi(j)}{\pi(l)}.$$

Given that we have landed in  $j$  for the first time, let  $\lambda_j$  be the probability, starting from  $j$ , that we return to the origin before another landing in  $j$ . This must occur with positive probability, otherwise  $P(S_n = 0 \text{ i.o.}) < 1$ . So whether or not another landing occurs before return is decided by tossing a coin with probability  $\lambda_j$  of failure. The expected number of additional landings past the first is given by the expected number of trials until failure. This is given by  $\sum_{m=1}^{\infty} m(1 - \lambda_j)^m \lambda_j < \infty$ , hence  $\pi(j)$  is finite. Add the convention  $\pi(0) = 1$ , then (3.46) holds for all states,  $j$  and  $l$ . Let  $n^*$  be the first time that state  $l$  is entered. By (3.46)

$$\frac{N_{n^*+n}(j)}{N_{n^*+n}(l)} \xrightarrow{\text{a.s.}} \frac{\pi(j)}{\pi(l)}.$$

But  $N_{n^*+n}(l)$  is the number of times that the sums  $S_{k+n^*} - S_{n^*}$ ,  $k = 1, \dots, n$  land in state zero, and  $N_{n^*+n}(j)$  is the number of times that  $S_{k+n^*} - S_{n^*}$ ,  $k = 1, \dots, n$  land in  $j - l$ , plus the number of landings in  $j$  by the sums  $S_k$  up to time  $n^*$ . Therefore,  $\pi(j)/\pi(l) = \pi(j - l)/\pi(0)$ , or

$$\pi(j) = \pi(l)\pi(j - l).$$

This is the exponential equation on the integers—the only solutions are  $\pi(j) = r^j$ . Consider any sequence of states  $m_1, \dots, m_{n-1}$ , 0 terminating at zero. I assert that

$$\begin{aligned} P(S_1 = m_1, \dots, S_{n-1} = m_{n-1}, S_n = 0) \\ = P(S_1 = -m_{n-1}, S_2 = -m_{n-2}, \dots, S_{n-1} = -m_1, S_n = 0). \end{aligned}$$

The first probability is  $P(X_1 = m_1, \dots, X_n = -m_{n-1})$ . The fact that  $X_1, \dots, X_n$  are identically distributed lets us equate this to

$$P(X_1 = -m_{n-1}, \dots, X_n = m_1),$$

which is the second probability. This implies that  $\pi(j) = \pi(-j)$ , hence  $\pi(j) \equiv 1$ .

**Problems**

19. For fair coin-tossing equalizations  $P(T_1 = 2n) = 1/n(2n-2)C_{n-1}2^{-2n+1}$ . Use this result to show that

$$a) P(T_1 = 2n) \sim \frac{1}{2\sqrt{\pi n}^{3/2}}.$$

b) From (a) conclude  $ET_1 = \infty$ .

c) Using (a) again, show that  $P(T_1 \geq 2n) \sim \frac{1}{\sqrt{\pi n}^{1/2}}$ .

d) Use (c) to show that  $P(T_k > 2k^2 \text{ i.o.}) = 1$ .

e) Conclude from (d) that  $P\left(\overline{\lim}_n \frac{R_n}{n} = \infty\right) = 1$ .

(There are a number of ways of deriving the exact expression above for  $P(T_1 = 2n)$ ; see Feller [59, Vol. I, pp. 74–75].)

20. For fair coin-tossing, use  $P(S_{2n} = 0) \sim 1/\sqrt{\pi n}$  to show that

$$E[N_n(0)]^2 \sim cn.$$

Use an argument similar to the proof of the strong law of large numbers for fair-coin tossing, Theorem 1.21, to prove

$$\frac{N_n(0)}{n} \xrightarrow{\text{a.s.}} 0.$$

21. Define

$$n_1^* = \min \{n; S_n > 0\},$$

$$n_2^* = \min \{n; S_{n+n_1^*} > S_{n_1^*}\}, \text{ and so forth,}$$

$$Y_k = \begin{cases} S_{n_1^*} + \cdots + n_k^* - S_{n_1^*} - \cdots - n_{k-1}^*, & k > 1, \\ S_{n_1^*}, & k = 1. \end{cases}$$

Show that the sequence  $(n_k^*, Y_k)$  are independent, identically distributed vectors. Use the law of large numbers to prove that if  $E|X_1| < \infty$ ,  $EX_1 > 0$ , then if one of  $EY_1$ ,  $En_1^*$  is finite, so is the other, and

$$(En_1^*)(EX_1) = EY_1.$$

Show by using the sequence

$$\tilde{X}_k = \begin{cases} X_k, & X_k \leq \lambda, \\ 0, & X_k > \lambda, \end{cases}$$

for  $\lambda > 0$ , that  $EY_1 < \infty$ ,  $En_1^* < \infty$ . (See Blackwell [7].)

22. For sums  $S_1, S_2, \dots$  such that  $P(S_n = 0 \text{ i.o.}) = 1$ , and  $R_1, R_2, \dots$  the occurrence times of  $\{S_n = 0\}$ , define the vectors  $Z_k$  by

$$Z_k = (S_{R_k+1}, \dots, S_{R_{k+1}}).$$

Define the appropriate range space  $(R, \mathcal{B})$  for each of these vectors and show that they are independent and identically distributed.

## 9. HEWITT-SAVAGE ZERO-ONE LAW

Section 7 proved that for any interval  $I$ ,  $P(S_n \in I \text{ i.o.})$  is zero or one. But these sets are not tail events; whether  $S_n = 0$  an infinite number of times depends strongly on  $X_1$ . However, there is another zero-one law in operation, formulated recently by Hewitt and Savage [71] which covers a variety of non-tail events.

**Definition 3.47.** For a process  $X_1, X_2, \dots$ ,  $A \in \mathcal{F}(X)$  is said to be symmetric if for any finite permutation  $\{i_1, i_2, \dots\}$  of  $\{1, 2, \dots\}$ , there is a set  $B \in \mathcal{B}_\infty$  such that

$$A = \{(X_1, X_2, \dots) \in B\} = \{(X_{i_1}, X_{i_2}, \dots) \in B\}.$$

**Theorem 3.48.** For  $X_1, X_2, \dots$  independent and identically distributed, every symmetric set has probability zero or one.

*Proof.* The short proof we give here is due to Feller [59, Vol. II]. Take  $A_n \in \mathcal{F}(X_1, \dots, X_n)$  so that  $P(A \triangle A_n) \rightarrow 0$ .  $A_n$  can be written

$$\{(X_1, \dots, X_n) \in B_n\}, \quad B_n \in \mathcal{B}_n.$$

Because the  $\tilde{X} = (X_{i_1}, X_{i_2}, \dots)$  process,  $i_1, i_2, \dots$  any sequence of distinct integers, has the same distribution as  $X$ ,

$$P(\tilde{X} \in C) = P(X \in C), \quad C \in \mathcal{B}_\infty.$$

Hence, for any  $B \in \mathcal{B}_\infty$ ,

$$(3.49) \quad P(\{\tilde{X} \in B\} \triangle \{X \in B_n\}) = P(\{X \in B\} \triangle \{X \in B_n\}).$$

Take  $(i_1, i_2, \dots) = (2n, 2n-1, \dots, 1, 2n+1, \dots)$ .

$$\tilde{A}_n = \{\tilde{X} \in B_n\} = \{(X_{2n}, \dots, X_{n+1}) \in B_n\}.$$

By (3.49), taking  $B \in \mathcal{B}_\infty$  such that  $\{\tilde{X} \in B\} = \{X \in B\} = A$ ,

$$P(A \triangle \tilde{A}_n) = P(A \triangle A_n) \rightarrow 0 \Rightarrow P(A_n \triangle \tilde{A}_n) \rightarrow 0 \Rightarrow P(A_n \cap \tilde{A}_n) \rightarrow P(A).$$

But  $A_n$  and  $\tilde{A}_n$  are independent, thus

$$P(A_n \cap \tilde{A}_n) = P(A_n)P(\tilde{A}_n) \rightarrow P(A)^2.$$

Again, as in the Kolmogorov zero-one law, we wind up with

$$P(A) = P(A)^2.$$

Q.E.D.

**Corollary 3.50.** For  $X_1, X_2, \dots$  independent and identically distributed,  $\{S_n\}$  the sums, every tail event on the  $S_1, S_2, \dots$  process has probability zero or one.

*Proof.* For  $A$  a tail event, if  $\{i_1, i_2, \dots\}$  permutes only the first  $n$  indices, take  $B \in \mathcal{B}_\infty$  such that

$$A = \{(S_{n+1}, \dots) \in B\}.$$

Thus  $A$  is a symmetric set.

This result leads to the mention of another famous strong limit theorem which will be proved much later. If  $EX_k = 0$ ,  $EX_k^2 < \infty$  for independent, identically distributed random variables, then the form 3.27 of the strong law of large numbers implies

$$\frac{X_1 + \dots + X_n}{\sqrt{n} \log n} \xrightarrow{\text{a.s.}} 0.$$

On the other hand, it is not hard to show that

$$\liminf \frac{X_1 + \dots + X_n}{\sqrt{n}} = -\infty \text{ a.s.}$$

(3.51)

$$\limsup \frac{X_1 + \dots + X_n}{\sqrt{n}} = +\infty \text{ a.s.}$$

Therefore fluctuations of  $S_n$  should be somewhere in the range  $\sqrt{n}$  to  $\sqrt{n} \log n$ . For any function  $h(n) \uparrow \infty$ , the random variable  $\overline{\lim} |S_n|/h(n)$  is a tail random variable, hence a.s. constant. The famous law of the iterated logarithm is

**Theorem 3.52**

$$\overline{\lim} \frac{|S_n|}{\sigma \sqrt{2n \log(\log n)}} = 1 \text{ a.s.}$$

This is equivalent to: For every  $\epsilon > 0$ ,

$$P(|S_n| \geq (1 - \epsilon)h(n) \text{ i.o.}) = 1,$$

$$P(|S_n| \geq (1 + \epsilon)h(n) \text{ i.o.}) = 0,$$

with  $h(n) = \sigma \sqrt{2n \log(\log n)}$ .

Therefore, a more general version of the law of the iterated logarithm would be a separation of all nondecreasing  $h(n)$  into two classes

$$P(|S_n| > h(n) \text{ i.o.}) = 0 \quad \text{or} \quad 1.$$

The latter dichotomy holds because of 3.50. The proof of 3.52 is quite tricky, to say nothing of the more general version. The simplest proof around for coin-tossing is in Feller [59, Vol. I].

Actually, this theorem is an oddity. Because, even though it is a strong limit theorem, it is a second-moment theorem and its proof consists of ingenious uses of the Borel-Cantelli lemma combined with the central limit theorem. We give an illuminating proof in Chapter 13.

**Problem 23.** Use the Kolmogorov zero-one law and the central limit theorem to prove (3.51) for fair coin-tossing.

*Remark.* The important theorems for independence come up over and over again as their contents are generalized. In particular, the random signs problem connects with martingales (Chapter 5), the strong law of large numbers generalizes into the ergodic theorem (Chapter 6), and the notions of recurrence of sums comes up again in Markov processes (Chapter 7).

## NOTES

The strong law of large numbers was proven for fair coin-tossing by Borel in 1909. The forms of the strong law given in this chapter were proved by Kolmogorov in 1930, 1933 [92], [98]. The general solution of the random signs problem is due to Khintchine and Kolmogorov [91] in 1925. A special case of the law of the iterated logarithm was proven by Khintchine [88, 1924].

The work on recurrence is more contemporary. The theorems of Section 7 are due to Chung and Fuchs [18, 1951], but the neat proof given that  $EX = 0$  implies recurrence was found by Chung and Ornstein [20, 1962]. But this work was preceded by some intriguing examples due to Polya [116, 1921]. (These will be given in Chapter 7). The work of Harris and Robbins (loc. cit.) on the equidistribution of sums appeared in 1953.

There is a bound for sums of independent random variables with zero means which is much more well-known than the Skorokhod's inequality, that is,

$$P\left(\max_{k \leq n} |S_k| > \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{k=1}^n EX_k^2,$$

$S_k = X_1 + \cdots + X_k$ . This is due to Kolmogorov. Compare it with the Chebyshev bound for  $P(|S_n| > \epsilon)$ . A generalization is proved in Chapter 5.

The strong law for identically distributed random variables depends essentially on  $E|X| < \infty$ . One might expect that even if  $E|X| = \infty$ , there would be another normalization  $N_n \uparrow \infty$  such that the normed sums

$$\frac{X_1 + \cdots + X_n}{N_n}$$

converge a.s. One answer is trivial; you can always take  $N_n$  increasing so rapidly that a.s. convergence to zero follows. But Chow and Robbins [14]

have obtained the strong result that if  $E|X| = \infty$ , there is no normalization  $N_n \uparrow \infty$  such that one gets a.s. convergence to anything but zero.

If  $E|X_1| < \infty$ , then if the sums are nonrecurrent, either  $S_n \rightarrow +\infty$  a.s. or  $S_n \rightarrow -\infty$  a.s. But if  $E|X_1| = \infty$ , the sums can be transient and still change sign an infinite number of times; in fact, one can get  $\overline{\lim} S_n = +\infty$  a.s.,  $\underline{\lim} S_n = -\infty$  a.s. Examples of this occur when  $X_1, X_2, \dots$  have one of the symmetric stable distributions discussed in Chapter 9.

Strassen [135] has shown that the law of the iterated logarithm is a second-moment theorem in the sense that if  $EX_1 = 0$ , and

$$\overline{\lim} \frac{X_1 + \cdots + X_n}{\sqrt{n \log(\log n)}} < \infty \text{ a.s.},$$

then  $EX_1^2 < \infty$ . There is some work on other forms of this law when  $EX_1^2 = \infty$ , but the results (Feller [54]) are very specialized.

For more extensive work with independent random variables the most interesting source remains Paul Lévy's book [103]. Loève's book has a good deal of the classical material. For an elegant and interesting development of the ideas of recurrence see Spitzer [130].

## CHAPTER 4

# CONDITIONAL PROBABILITY AND CONDITIONAL EXPECTATION

### 1. INTRODUCTION

More general tools need to be developed to handle relationships between dependent random variables. The concept of conditional probability—the distribution of one set of random variables given information concerning the observed values of another set—will turn out to be a most useful tool.

First consider the problem: What is the probability of an event  $B$  given that  $A$  has occurred? If we know that  $\omega \in A$ , then our new sample space is  $A$ . The probability of  $B$  is proportional to the probability of that part of it lying in  $A$ . Hence

**Definition 4.1.** Given  $(\Omega, \mathcal{F}, P)$ , for sets  $A, B \in \mathcal{F}$ , such that  $P(A) > 0$ , the conditional probability of  $B$  given that  $A$  has occurred is defined as

$$\frac{P(B \cap A)}{P(A)},$$

and is denoted by  $P(B | A)$ .

This extends immediately to conditioning by random variables taking only a countable number of values.

**Definition 4.2.** If  $X$  takes values in  $\{x_k\}$ , the conditional probability of  $A$  given  $X = x_k$  is defined by

$$P(A | X = x_k) = \frac{P(A, X = x_k)}{P(X = x_k)}$$

if  $P(X = x_k) > 0$  and arbitrarily defined as zero if  $P(X = x_k) = 0$ .

Note that there is probability zero that  $X$  takes values in the set where the conditional probability was not defined by the ratio.  $P(A | X = x_k)$  is a probability on  $\mathcal{F}$ , and the natural definition of the conditional expectation of a random variable  $Y$  given  $X = x_k$  is

$$(4.3) \quad E(Y | X = x_k) = \int Y(\omega) P(d\omega | X = x_k)$$

if the integral exists.



What needs to be done is to generalize the definition so as to be able to handle random variables taking on nondenumerably many values. Look at the simplest case of this: Suppose there is a random variable  $X$  on  $(\Omega, \mathcal{F}, P)$ , and let  $A \in \mathcal{F}$ . If  $B \in \mathcal{B}_1$  is such that  $P(X \in B) > 0$ , then as above, the conditional probability of  $A$  given  $X \in B$ , is defined by

$$P(A | X \in B) = \frac{P(A, X \in B)}{P(X \in B)}.$$

But suppose we want to give meaning to the conditional probability of  $A$  given  $X(\omega) = x$ . Of course, if  $P(X = x) > 0$ , then we have no trouble and proceed as in 4.2. But many of the interesting random variables have the property that  $P(X = x) = 0$  for all  $x$ . This causes a fuss. An obvious thing to try is taking limits, i.e., to try defining

$$(4.4) \quad P(A | X = x) = \lim_{h \downarrow 0} \frac{P(A, X \in (x - h, x + h))}{P(X \in (x - h, x + h))}.$$

In general, this is no good. If  $P(X = x_0) = 0$ , then there is no guarantee, unless we put more restrictive conditions on  $P$  and  $X$ , that the limit above will exist for

$$x = x_0.$$

So either we add these restrictions (very unpleasant), or we look at the problem a different way. Look at the limit in (4.4) globally as a function of  $x$ . Intuitively, it looks as though we are trying to take the derivative of one measure with respect to another. This has a familiar ring; we look back to see what can be done.

On  $\mathcal{B}_1$  define two measures as follows: Let

$$(4.5) \quad \begin{aligned} \hat{Q}(B) &= P(A, X \in B), \\ \hat{P}(B) &= P(X \in B). \end{aligned}$$

Note that  $0 \leq \hat{Q}(B) \leq \hat{P}(B)$  so that  $\hat{Q}$  is absolutely continuous with respect to  $\hat{P}$ . By the Radon-Nikodym theorem (Appendix A.30) we can define the derivative of  $\hat{Q}$  with respect to  $\hat{P}$ , which is exactly what we are trying to do with limits in (4.4). But we must pay a price for taking this elegant route. Namely, recall that  $d\hat{Q}/d\hat{P}$  is defined as any  $\mathcal{B}_1$ -measurable function  $\varphi(x)$  satisfying

$$(4.6) \quad \hat{Q}(B) = \int_B \varphi(x) \hat{P}(dx), \quad \text{all } B \in \mathcal{B}_1.$$

If  $\varphi$  satisfies (4.6) so does  $\varphi'$  if  $\varphi = \varphi'$  a.s.  $\hat{P}$ . Hence this approach, defining  $P(A | X = x)$  as any function satisfying (4.6) leads to an arbitrary selection of one function from among a class of functions equivalent (a.s. equal) under  $\hat{P}$ . This is a lesser evil.

**Definition 4.7.** The conditional probability  $P(A | \mathbf{X} = x)$  is defined as any  $\mathcal{B}_1$ -measurable function satisfying

$$P(A, \mathbf{X} \in B) = \int_B P(A | \mathbf{X} = x) \hat{P}(dx), \quad \text{all } B \in \mathcal{B}_1.$$

In 4.7 above,  $P(A | \mathbf{X} = x)$  is defined as a  $\mathcal{B}_1$ -measurable function  $\varphi(x)$ , unique up to equivalence under  $\hat{P}$ . For many purposes it is useful to consider the conditional probability as a random variable on the original  $(\Omega, \mathcal{F}, P)$  space, rather than the version above which resembles going into representation space. The natural way to do this is to define

$$P(A | \mathbf{X}(\omega)) = \varphi(\mathbf{X}(\omega)).$$

Since  $\varphi$  is  $\mathcal{B}_1$ -measurable, then  $\varphi(\mathbf{X}(\omega))$  is a random variable on  $(\Omega, \mathcal{F})$ . Since any two versions of  $\varphi$  are equivalent under  $\hat{P}$ , any two versions of  $P(A | \mathbf{X}(\omega))$  obtained in this way are equivalent under  $P$ . But there is a more direct way to get to  $P(A | \mathbf{X})$ , analogous to 4.7. Actually, what is done is just transform 4.7 to  $(\Omega, \mathcal{F}, P)$ .

**Definition 4.8.** The conditional probability of  $A$  given  $\mathbf{X}(\omega)$ , is defined as any random variable on  $\Omega$ , measurable  $\mathcal{F}(\mathbf{X})$ , and satisfying

$$P(A, \mathbf{X} \in B) = \int_{\{\mathbf{X} \in B\}} P(A | \mathbf{X}) dP, \quad \text{all } B \in \mathcal{B}_1.$$

Any two versions of  $P(A | \mathbf{X})$  differ on a set of probability zero.

This gives the same  $P(A | \mathbf{X})$  as starting from 4.7 to get  $\varphi(\mathbf{X}(\omega))$ , where  $\varphi(x) = P(A | \mathbf{X} = x)$ . To see this, apply 2.41 to 4.7 and compare the result with 4.8. A proof that is a bit more interesting utilizes a converse of 2.31.

**Proposition 4.9.** Let  $\mathbf{X}$  be a random vector on  $(\Omega, \mathcal{F})$  taking values in  $(R, \mathcal{B})$ . If  $Z$  is a random variable on  $(\Omega, \mathcal{F})$ , measurable  $\mathcal{F}(\mathbf{X})$ , then there is a random variable  $\theta(\mathbf{x})$  on  $(R, \mathcal{B})$  such that

$$Z = \theta(\mathbf{X}).$$

*Proof.* See Appendix A.21.

The fact that  $P(A | \mathbf{X})$  is  $\mathcal{F}(\mathbf{X})$ -measurable implies by this proposition that  $P(A | \mathbf{X}) = \theta(\mathbf{X})$ , where  $\theta(x)$  is  $\mathcal{B}_1$ -measurable. But  $\theta(\mathbf{X})$  satisfies

$$P(A, \mathbf{X} \in B) = \int_{\{\mathbf{X} \in B\}} \theta(\mathbf{X}(\omega)) dP = \int_B \theta(x) \hat{P}(dx)$$

(this last by 2.41). Hence  $\theta = \varphi$  a.s.  $\hat{P}$ .

We can put 4.8 into a form which shows up a seemingly curious phenomenon. Since  $\mathcal{F}(\mathbf{X})$  is the class of all sets  $\{\mathbf{X} \in B\}$ ,  $B \in \mathcal{B}_1$ ,  $P(A | \mathbf{X})$  is any

random variable satisfying

$$(4.10) \quad P(A \cap D) = \int_D P(A | X) dP, \quad \text{all } D \in \mathcal{F}(X).$$

From this, make the observation that if  $X_1$  and  $X_2$  are two random variables which contain the same information in the sense that  $\mathcal{F}(X_1) = \mathcal{F}(X_2)$ , then

$$(4.11) \quad P(A | X_1) = P(A | X_2) \quad \text{a.s.}$$

In a way this is not surprising, because  $\mathcal{F}(X_1) = \mathcal{F}(X_2)$  implies that  $X_1$  and  $X_2$  are functions of each other, that is, from 4.9,

$$X_2(\omega) = \theta_1(X_1(\omega)), \quad X_1(\omega) = \theta_2(X_2(\omega)).$$

The idea here is that  $P(A | X)$  does not depend on the values of  $X$ , but rather on the sets in  $\mathcal{F}$  that  $X$  discriminates between.

The same course can be followed in defining the conditional expectation of one random variable, given the value of another. Let  $X, Y$  be random variables on  $(\Omega, \mathcal{F}, P)$ . What we wish to define is the conditional expectation of  $Y$  given  $X = x$ , in symbols,  $E(Y | X = x)$ . If  $B \in \mathcal{B}$  were such that  $P(X \in B) > 0$ , intuitively  $E(Y | X \in B)$  should be defined as  $\int Y(\omega)P(d\omega | X \in B)$ , where  $P(\cdot | X \in B)$  is the probability on  $\mathcal{F}$  defined as

$$P(A | X \in B) = P(A, X \in B)/P(X \in B).$$

Again, we could take  $B = (x - h, x + h)$ , let  $h \downarrow 0$ , and hope the limit exists. More explicitly, we write the ratio

$$\frac{\int Y(\omega)P(d\omega, X \in B)}{P(X \in B)}$$

and hope that as  $P(X \in B) \rightarrow 0$ , the limiting ratio exists. Again the derivative of one set function with respect to another is coming up. What to do is similar: Define

$$\hat{Q}(B) = \int Y(\omega)P(d\omega, X \in B),$$

$$\hat{P}(B) = P(X \in B).$$

To get things finite, we have to assume  $E|Y| < \infty$ ; then

$$|\hat{Q}(B)| \leq \int |Y| P(d\omega, X \in B) \leq E|Y|.$$

To show that  $\hat{Q}$  is  $\sigma$ -additive, write it as

$$\hat{Q}(B) = \int_{\{X \in B\}} Y(\omega) dP.$$

Now  $\{B_n\}$  disjoint implies  $A_n = \{X \in B_n\}$  disjoint, and

$$\hat{Q}(\cup B_n) = \int_{\cup A_n} Y dP = \sum_n \int_{A_n} Y dP = \sum_n \hat{Q}(B_n).$$

Also,  $\hat{P}(B) = 0 \Rightarrow \hat{Q}(B) = 0$ , thus  $\hat{Q}$  is absolutely continuous with respect to  $\hat{P}$ . This allows the definition of  $E(Y | X = x)$  as any version of  $d\hat{Q}/d\hat{P}$ .

**Definition 4.12.** Let  $E|Y| < \infty$ , then  $E(Y | X = x)$  is any  $\mathcal{B}_1$ -measurable function satisfying

$$\int_B E(Y | X = x) d\hat{P}(x) = \int_{\{X \in B\}} Y(\omega) dP(\omega), \quad \text{all } B \in \mathcal{B}_1.$$

Any two versions of  $E(Y | X = x)$  are a.s. equal under  $\hat{P}$ .

Conditional expectations can also be looked at as random variables. Just as before, if  $\varphi(x) = E(Y | X = x)$ ,  $E(Y | X)$  can be defined as  $\varphi(X(\omega))$ . Again, we prefer the direct definition.

**Definition 4.13.** Let  $E|Y| < \infty$ ; then  $E(Y | X)$  is any  $\mathcal{F}(X)$ -measurable function satisfying

$$(4.14) \quad \int_A E(Y | X) dP = \int_A Y dP, \quad \text{all } A \in \mathcal{F}(X).$$

The random variable  $Y$  trivially satisfies (4.14), but in general  $Y \neq E(Y | X)$  because  $Y(\omega)$  is not necessarily  $\mathcal{F}(X)$ -measurable. This remark does discover the property that if  $\mathcal{F}(Y) \subset \mathcal{F}(X)$  then  $E(Y | X) = Y$  a.s. Another property in this direction is: Consider the space of  $\mathcal{F}(X)$ -measurable random variables. In this space, the random variable closest to  $Y$  is  $E(Y | X)$ . (For a defined version of this statement see Problem 11.)

Curiously enough, conditional probabilities are a special case of conditional expectations. Because, by the definitions,

$$\begin{aligned} P(A | X = x) &= E(\chi_A(\omega) | X = x) \quad \text{a.s. } \hat{P}, \\ P(A | X) &= E(\chi_A(\omega) | X) \quad \text{a.s. } P. \end{aligned}$$

Therefore, the next section deals with the general definition of conditional expectation.

**Definition 4.15.** Random variables  $X, Y$  on  $(\Omega, \mathcal{F}, P)$  are said to have a joint density if the probability  $\hat{P}(\cdot)$  defined on  $\mathcal{B}_2$  by  $\hat{P}(F) = P((Y, X) \in F)$  is absolutely continuous with respect to Lebesgue measure on  $\mathcal{B}_2$ , that is, if there exists  $f(y, x)$  on  $R^{(2)}$ , measurable  $\mathcal{B}_2$ , such that

$$P((Y, X) \in F) = \int_F f(y, x) dy dx, \quad \text{all } F \in \mathcal{B}_2.$$

Then, by Fubini's theorem (see Appendix A.37), defining  $f(x) = \int f(y, x) dy$ , for all  $B \in \mathcal{B}_1$ ,

$$\hat{P}(B) = P(X \in B) = \int_B f(x) dx.$$

(Actually, any  $\sigma$ -finite product measure on  $\mathcal{B}_2$  could be used instead of  $dy dx$ .) If a joint density exists, then it can be used to compute the conditional probability and expectation. This is the point of Problems 2 and 3 below.

### Problems

1. Let  $X$  take on only integer values. Show that  $P(A | X = x)$  as defined in 4.7 is any  $\mathcal{B}_1$ -measurable function  $\varphi(x)$  satisfying

$$P(A, X = j) = \varphi(j)P(X = j), \quad \text{all } j.$$

Conclude that if  $P(X = j) > 0$ , then any version of the above conditional probability satisfies

$$P(A | X = j) = \frac{P(A, X = j)}{P(X = j)}.$$

2. Prove that if  $X, Y$  have a joint density, then for any  $B \in \mathcal{B}_1$ ,

$$P(Y \in B | X = x) = \int_B \frac{f(y, x)}{f(x)} dy, \quad \text{a.s. } \hat{P}.$$

3. If  $(Y, X)$  have a joint density  $f(y, x)$  and  $E|Y| < \infty$ , show that one version of  $E(Y | X = x)$  is given by

$$\int y \frac{f(y, x)}{f(x)} dy.$$

4. Let  $X_1, X_2$  take values in  $\{1, 2, \dots, N\}$ . If  $\mathcal{F}(X_1) = \mathcal{F}(X_2)$ , then prove there is a permutation  $\{i_1, i_2, \dots, i_N\}$  of  $\{1, 2, \dots, N\}$  such that

$$A_j = \{X_1 = j\} = \{X_2 = i_j\}.$$

Let  $P(A_j) > 0, j = 1, \dots, N$ . Prove that

$$P(A | X_1) = P(A | X_2) = \sum_1^N \chi_{A_j}(\omega) \cdot P(A | A_j) \quad \text{a.s.}$$

5. Let  $\Omega = \{z \in [-1, +1]\}$ ,  $\mathcal{F} = \mathcal{B}([-1, +1])$ ,  $P = dz/2$ , and  $X_1(z) = z^2$ . Show that one version of  $P(A | X_1)$  is

$$P(A | X_1) = \frac{1}{2}\chi_A(z) + \frac{1}{2}\chi_A(-z).$$

Let  $X_2(z) = z^4$ ; find a version of  $P(A | X_2)$ . Find versions of  $P(A | X_1 = x)$ ,  $P(A | X_2 = x)$ .

6. Given the situation of Problem 5, and with  $Y$  a random variable such that  $E|Y(z)| < \infty$ , show that

$$E(Y(z) | X_1(z)) = \frac{1}{2}Y(z) + \frac{1}{2}Y(-z), \quad \text{a.s. } dz.$$

Find a version of

$$E(Y(z) | X_1(z) = x).$$

7. If  $X$  and  $Y$  are independent, show that for any  $B \in \mathcal{B}_1$ ,  $P(Y \in B | X) = P(Y \in B)$  a.s. For  $E|Y| < \infty$ , show that  $E(Y | X) = EY$  a.s.

8. Give an example to show that  $E(Y | X) = EY$  a.s. does not imply that  $X$  and  $Y$  are independent.

9. (Borel paradox). Take  $\Omega$  to be the unit sphere  $S^{(2)}$  in  $R^{(3)}$ ,  $\mathcal{F}$  the Borel subsets of  $\Omega$ ,  $P(\cdot)$  the extension of surface area. Choose two opposing points on  $S^{(2)}$  as the poles and fix a reference half-plane passing through them. For any point  $p$ , define its longitude  $\psi(p)$  as the angle between  $-\pi$  and  $\pi$  that the half-plane of the great semi-circle through  $p$  makes with the reference half plane. Define its latitude  $\theta(p)$  as the angle that the radius to  $p$  makes with the equatorial plane,  $-\pi/2 < \theta(p) \leq \pi/2$ . Prove that the conditional probability of  $\psi$  given  $\theta$  is uniformly distributed over  $[-\pi, \pi)$  but that the conditional probability of  $\theta$  given  $\psi$  is not uniformly distributed over  $(-\pi/2, \pi/2]$ . (See Kolmogorov [98, p. 50].)

## 2. A MORE GENERAL CONDITIONAL EXPECTATION

Section 1 pointed out that  $E(Y | X)$  or  $P(A | X)$  depended only on  $\mathcal{F}(X)$ . The point was that the relevant information contained in knowing  $X(\omega)$  is the information regarding the location of  $\omega$ . Let  $(\Omega, \mathcal{F}, P)$  be a probability space,  $Y$  a random variable,  $E|Y| < \infty$ ,  $\mathcal{D}$  any  $\sigma$ -field,  $\mathcal{D} \subset \mathcal{F}$ .

**Definition 4.16.** The conditional expectation  $E(Y | \mathcal{D})$  is any random variable measurable  $(\Omega, \mathcal{D})$  such that

$$(4.17) \quad \int_D E(Y | \mathcal{D}) dP = \int_D Y dP, \quad \text{all } D \in \mathcal{D}.$$

As before, any two versions differ on a set of probability zero. If  $\mathcal{D} = \mathcal{F}(X_1, X_2, \dots)$ , denote  $E(Y | \mathcal{D}) = E(Y | X_1, X_2, \dots)$ .

If  $X$  is a random vector to  $(R, \mathcal{B})$ , then for  $x \in R$

**Definition 4.18.**  $E(Y | X = x)$  is any random variable on  $(R, \mathcal{B})$ , where  $\hat{P}(B) = P(X \in B)$ , satisfying

$$\int_B E(Y | X = x) d\hat{P} = \int_{\{X \in B\}} Y dP, \quad \text{all } B \in \mathcal{B}.$$

The importance of this is mostly computational. By inspection verify that

$$(4.19) \quad \varphi(\mathbf{x}) = E(Y | \mathbf{X} = \mathbf{x}) \Rightarrow E(Y | \mathbf{X}) = \varphi(\mathbf{X}).$$

Usually,  $E(Y | \mathbf{X} = \mathbf{x})$  is easier to compute, when densities exist, for example. Then (4.19) gets us  $E(Y | \mathbf{X})$ .

**Proposition 4.20.** *A list of properties of  $E(Y | \mathcal{D})$ ,*

- 1)  $E(\alpha Y_1 + \beta Y_2 | \mathcal{D}) = \alpha E(Y_1 | \mathcal{D}) + \beta E(Y_2 | \mathcal{D})$  a.s., assuming  $E|Y_1| < \infty$ ,  $E|Y_2| < \infty$ .
- 2)  $Y \geq 0, EY < \infty \Rightarrow E(Y | \mathcal{D}) \geq 0$  a.s.
- 3)  $E|Y| < \infty, \mathcal{D} \subset \mathcal{E}$ , implies

$$E(E(Y | \mathcal{E}) | \mathcal{D}) = E(Y | \mathcal{D}) \quad \text{a.s.}$$

- 4)  $\mathcal{F}(Y)$  independent of  $\mathcal{D}$ ,  $E|Y| < \infty \Rightarrow E(Y | \mathcal{D}) = EY$  a.s.

*Proofs.* These proofs follow pretty trivially from the definition 4.16. To improve technique I'll briefly go through them; the idea in all cases is to show that the integrals of both sides of (1), (2), (3), (4) over  $\mathcal{D}$  sets are the same, (in 2,  $\geq$ ). Let  $D \in \mathcal{D}$ . Then by 4.16 the integrals over  $D$  of the left hand sides of (1), (2), (3), (4) above are

$$\begin{array}{ll} 1) \int_D (\alpha Y_1 + \beta Y_2) dP & 2) \int_D Y dP \\ 3) \int_D E(Y | \mathcal{E}) dP & 4) \int_D Y dP. \end{array}$$

The right-hand sides integrated over  $D$  are

$$\begin{array}{ll} 1') \alpha \int_D Y_1 dP + \beta \int_D Y_2 dP & 2') 0 \\ 3') \int_D Y dP & 4') (EY) \cdot P(D) \end{array}$$

So (1) = (1') is trivial, also (2)  $\geq$  (2'). For (4), write

$$\int_D Y dP = \int \chi_D Y dP = E\chi_D EY = P(D) \cdot EY.$$

For (3), (3') note that by 4.16,

$$\int_F E(Y | \mathcal{E}) dP = \int_F Y dP, \quad \text{all } F \in \mathcal{E}.$$

But  $\mathcal{D} \subset \mathcal{E}$ , so (3) = (3'), all  $D \in \mathcal{D}$ .

An important property of the conditional expectation is, if  $E|Y| < \infty$ ,

$$(4.21) \quad E(E(Y | \mathcal{D})) = EY.$$

This follows quickly from the definitions.

Let  $Y = \chi_A(\omega)$ ; then the general definition of conditional probability is

**Definition 4.22.** Let  $\mathcal{D}$  be a sub- $\sigma$ -field of  $\mathcal{F}$ . The conditional probability of  $A \in \mathcal{F}$  given  $\mathcal{D}$  is a random variable  $P(A | \mathcal{D})$  on  $(\Omega, \mathcal{D})$  satisfying

$$\int_D P(A | \mathcal{D}) dP = P(A \cap D), \quad \text{all } D \in \mathcal{D}.$$

By the properties in 4.20, a conditional probability acts almost like a probability, that is,

$$(4.23) \quad P(A | \mathcal{D}) \geq 0 \quad \text{a.s.}, \quad P(\Omega | \mathcal{D}) = 1 \quad \text{a.s.}$$

$$A_1, \dots, A_n \text{ disjoint} \Rightarrow P(\cup A_n | \mathcal{D}) = \sum_1^n P(A_k | \mathcal{D}) \quad \text{a.s.}$$

It is also  $\sigma$ -additive almost surely. This follows from

**Proposition 4.24.** Let  $Y_n \geq 0$  be random variable such that  $Y_n \uparrow Y$  a.s. and  $E|Y| < \infty$ . Then  $E(Y_n | \mathcal{D}) \rightarrow E(Y | \mathcal{D})$  a.s.

*Proof.* Let  $Z_n = Y - Y_n$ , so  $Z_n \downarrow 0$  a.s., and  $EZ_n \downarrow 0$ . By 4.20(2)  $Z_n \geq Z_{n+1} \Rightarrow E(Z_n | \mathcal{D}) \geq E(Z_{n+1} | \mathcal{D})$  a.s. Therefore the sequence  $E(Z_n | \mathcal{D})$  converges monotonically downward a.s. to a random variable  $U \geq 0$ . By the monotone convergence theorem,

$$\lim_n E(E(Z_n | \mathcal{D})) = EU.$$

Equation (4.21) now gives  $EU = \lim_n EZ_n = 0$ . Thus  $U = 0$  a.s.

Let  $A_k \in \mathcal{F}$ ,  $\{A_k\}$  disjoint, and take

$$Y_n = \sum_{k=1}^n \chi_{A_k}$$

in the above proposition to get from (4.23) to

$$(4.25) \quad P(\cup A_n | \mathcal{D}) = \sum_n P(A_n | \mathcal{D}) \quad \text{a.s.}$$

For  $A$  fixed,  $P(A | \mathcal{D})$  is an equivalence class of functions  $f(A, \omega)$ . It seems reasonable to hope from (4.25) that from each equivalence class a function  $f^*(A, \omega)$  could be selected such that the resulting function  $P^*(A | \mathcal{D})$  on  $\mathcal{F} \times \Omega$  would be a probability on  $\mathcal{F}$  for every  $\omega$ . If this can be done, then the entire business of defining  $E(Y | \mathcal{D})$  would be unnecessary because it



could be defined as

$$(4.26) \quad E(Y | \mathcal{D}) = \int Y(\omega) P^*(d\omega | \mathcal{D}).$$

Unfortunately, in general it is not possible to do this. What can be done is a question which is formulated and partially answered in the next section.

### Problems

**10.** If  $\mathcal{D}$  has the property,  $D \in \mathcal{D} \Rightarrow P(D) = 0, 1$ , then show  $E(Y | \mathcal{D}) = EY$  a.s. (if  $E|Y| < \infty$ ).

**11.** Let  $Y$  be a random variable on  $(\Omega, \mathcal{F})$ ,  $EY^2 < \infty$ . For any random variable  $X$  on  $(\Omega, \mathcal{F})$ , let  $d^2(Y, X) = E|Y - X|^2$ .

a) Prove that among all random variables  $X$  on  $(\Omega, \mathcal{D})$ ,  $\mathcal{D} \subset \mathcal{F}$ , there is an a.s. unique random variable  $Y_0$  which minimizes  $d(Y, X)$ . This random variable  $Y_0$  is called the best predictor of  $Y$  based on  $\mathcal{D}$ .

b) Prove that  $Y_0 = E(Y | \mathcal{D})$  a.s.

**12.** Let  $X_0, X_1, X_2, \dots, X_n$  be random variables having a joint normal distribution,  $EX_k = 0$ ,  $\Gamma_{ij} = E(X_i X_j)$ . Show that  $E(X_0 | X_1, X_2, \dots, X_n) = \sum_{j=1}^n \lambda_j X_j$  a.s., and give the equations determining  $\lambda_j$  in terms of the  $\Gamma_{ij}$ . (See Chapter 9 for definitions.)

**13.** Let  $X$  be a random vector taking values in  $(R, \mathcal{B})$ , and  $\varphi(x)$  a random variable on  $(R, \mathcal{B})$ . Let  $Y$  be a random variable on  $(\Omega, \mathcal{F})$  and  $E|Y| < \infty$ ,  $E|\varphi(X)| < \infty$ ,  $E|Y\varphi(X)| < \infty$ . Prove that

a)  $E(\varphi(X)Y | X) = \varphi(X)E(Y | X)$  a.s.  $P$ .

b)  $E(\varphi(X)Y | X = x) = \varphi(x)E(Y | X = x)$  a.s.  $\hat{P}$ .

[This result concurs with the idea that if we know  $X$ , then given  $X$ ,  $\varphi(X)$  should be treated as a constant. To work Problem 13, a word to the wise: Start by assuming  $\varphi(X) \geq 0$ ,  $Y \geq 0$ , consider the class of random variables  $\varphi$  for which it is true, and apply 2.38.]

**14.** Let  $Y$  be a random variable on  $(\Omega, \mathcal{F}, P)$  such that  $E|Y| < \infty$  and  $X_1, X_2$  random vectors such that  $\mathcal{F}(Y, X_1)$  is independent of  $\mathcal{F}(X_2)$ , then prove that

$$E(Y | X_1, X_2) = E(Y | X_1) \quad \text{a.s.}$$

**15.** Let  $X_1, X_2, \dots$  be independent, identically distributed random variables,  $E|X_1| < \infty$ , and denote  $S_n = X_1 + \dots + X_n$ . Prove that

$$E(X_1 | S_n, S_{n+1}, \dots) = S_n/n \quad \text{a.s.}$$

[Use symmetry in the final step.]

### 3. REGULAR CONDITIONAL PROBABILITIES AND DISTRIBUTIONS

**Definition 4.27.**  $P^*(A | \mathcal{D})$  will be called a regular conditional probability on  $\mathcal{F}_1 \subset \mathcal{F}$ , given  $\mathcal{D}$  if

- a) For  $A \in \mathcal{F}_1$  fixed,  $P^*(A | \mathcal{D})$  is a version of  $P(A | \mathcal{D})$ .
- b) For any  $\omega$  held fixed,  $P^*(A | \mathcal{D})$  is a probability on  $\mathcal{F}_1$ .

If a regular conditional probability, given  $\mathcal{D}$ , exists, all the conditional expectations can be defined through it.

**Proposition 4.28.** If  $P^*(A | \mathcal{D})$  is a regular conditional probability on  $\mathcal{F}_1$  and if  $Y$  is a random variable on  $(\Omega, \mathcal{F}_1)$ ,  $E|Y| < \infty$ , then

$$E(Y | \mathcal{D}) = \int Y(\omega) P^*(d\omega | \mathcal{D}) \quad \text{a.s.}$$

*Proof.* Consider all nonnegative random variables on  $(\Omega, \mathcal{F}_1)$  for which 4.28 holds. For the random variable  $\chi_A$ ,  $A \in \mathcal{F}_1$ , 4.28 becomes

$$P(A | \mathcal{D}) = P^*(A | \mathcal{D}) \quad \text{a.s.,}$$

which holds by 4.27. Hence 4.28 is true for simple functions. Now for  $Y \geq 0$ ,  $EY < \infty$  take  $Y_n$  simple  $\uparrow Y$ . Then by 4.24 and the monotone convergence theorem,

$$E(Y | \mathcal{D}) = \lim_n E(Y_n | \mathcal{D}) = \lim_n \int Y_n dP^*(d\omega | \mathcal{D}) = \int Y P^*(d\omega | \mathcal{D}) \quad \text{a.s.}$$

Unfortunately, a regular conditional probability on  $\mathcal{F}_1$  given  $\mathcal{D}$  does not exist in general (see the Chapter notes). The difficulty is this: by (4.25), for  $A_n \in \mathcal{F}_1$ , disjoint, there is a set of probability zero such that

$$P(\cup A_n | \mathcal{D}) \neq \sum_n P(A_n | \mathcal{D}),$$

If  $\mathcal{F}_1$  contains enough countable collections of disjoint sets, then the exceptional sets may pile up.

By doing something which is like passing over to representation space we can get rid of this difficulty. Let  $\mathbf{Y}$  be a random vector taking values in a space  $(R, \mathcal{B})$ .

**Definition 4.29.**  $\hat{P}(B | \mathcal{D})$  defined for  $B \in \mathcal{B}$  and  $\omega \in \Omega$  is called a regular conditional distribution for  $\mathbf{Y}$  given  $\mathcal{D}$  if

- i) for  $B \in \mathcal{B}$  fixed,  $\hat{P}(B | \mathcal{D})$  is a version of  $P(\mathbf{Y} \in B | \mathcal{D})$ .
- ii) for any  $\omega \in \Omega$  fixed,  $\hat{P}(B | \mathcal{D})$  is a probability on  $\mathcal{B}$ .

If  $\mathbf{Y}$  is a random variable, then by using the structure of  $R^{(1)}$  in an essential way we prove the following theorem.

**Theorem 4.30.** *There always exists a regular conditional distribution for a random variable  $Y$  given  $\mathcal{D}$ .*

*Proof.* The general proof we break up into steps.

**Definition 4.31.**  $F(x | \mathcal{D})$  on  $R^{(1)} \times \Omega$  is a conditional distribution function for  $Y$  given  $\mathcal{D}$  if

- i)  $F(x | \mathcal{D})$  is a version of  $P(Y < x | \mathcal{D})$  for every  $x$ .
- ii) for every  $\omega$ ,  $F(x | \mathcal{D})$  is a distribution function.

**Proposition 4.32.** *There always exists a conditional distribution function for  $Y$  given  $\mathcal{D}$ .*

*Proof.* Let  $R = \{r_j\}$  be the rationals. Select versions of  $P(Y < r_j | \mathcal{D})$  and define

$$M_{ij} = \{\omega; P(Y < r_j | \mathcal{D}) < P(Y < r_i | \mathcal{D})\},$$

$$M = \bigcup_{r_j > r_i} M_{ij}.$$

So  $M$  is the set on which monotonicity is violated. By 4.20(2)  $P(M) = 0$ . Define

$$N_i = \left\{ \omega; \lim_{r_j \uparrow r_i} P(Y < r_j | \mathcal{D}) \neq P(Y < r_i | \mathcal{D}) \right\}, \quad N = \bigcup_i N_i.$$

Since  $\chi_{(-\infty, r_j)}(Y) \uparrow \chi_{(-\infty, r_i)}(Y)$  4.24 implies  $P(N_i) = 0$ , or  $P(N) = 0$ . Finally take  $r_j \uparrow \infty$  or  $r_j \downarrow -\infty$  and observe that the set  $L$  on which  $\lim_{r_j} P(Y < r_j | \mathcal{D})$  fails to equal one or zero, respectively, has probability zero. Thus, for  $\omega$  in the complement of  $M \cup N \cup L$ ,  $P(Y < x | \mathcal{D})$  for  $x \in R$  is monotone, left-continuous, zero and one at  $-\infty, +\infty$ , respectively. Take  $G(x)$  an arbitrary distribution function and define:

$$F(x | \mathcal{D}) = \begin{cases} G(x), & \omega \in M \cup N \cup L, \\ \lim_{r_j \uparrow x} P(Y < r_j | \mathcal{D}), & \text{otherwise.} \end{cases}$$

It is a routine job to verify that  $F(x | \mathcal{D})$  defined this way is a distribution function for every  $\omega$ . Use 4.24 again and  $\chi_{(-\infty, r_j)}(Y) \uparrow \chi_{(-\infty, x)}(Y)$  to check that  $F(x | \mathcal{D})$  is a version of  $P(Y < x | \mathcal{D})$ .

Back to Theorem 4.30. For  $Y$  a random variable, define  $\hat{P}(\cdot | \mathcal{D})$ , for each  $\omega$ , as the probability on  $(R^{(1)}, \mathcal{B}_1)$  extended from the distribution function  $F(x | \mathcal{D})$ . Let  $\mathcal{C}$  be the class of all sets  $C \in \mathcal{B}_1$  such that  $\hat{P}(C | \mathcal{D})$  is a version of  $P(Y \in C | \mathcal{D})$ . By 4.23,  $\mathcal{C}$  contains all finite disjoint unions of left-closed right-open intervals. By 4.24,  $\mathcal{C}$  is closed under monotone limits. Hence  $\mathcal{C} = \mathcal{B}_1$ . Therefore,  $\hat{P}(\cdot | \mathcal{D})$  is the required conditional distribution.

For random vectors  $\mathbf{Y}$  taking values in  $(R, \mathcal{B})$ , this result can be extended if  $(R, \mathcal{B})$  looks enough like  $(R^{(1)}, \mathcal{B}_1)$ .

**Definition 4.33.** Call  $(R, \mathcal{B})$  a Borel space if there is a  $E \in \mathcal{B}_1$  and a one-to-one mapping  $\varphi: R \leftrightarrow E$  such that  $\varphi$  is  $\mathcal{B}$ -measurable and  $\varphi^{-1}$  is  $\mathcal{B}_1$ -measurable.

Borel spaces include almost all the useful probability spaces. For example, see the proof in Appendix A that  $(R^{(\infty)}, \mathcal{B}_\infty)$  is a Borel space. So, more easily, is  $(R^{(n)}, \mathcal{B}_n)$ ,  $n \geq 1$ .

**Theorem 4.34.** If  $\mathbf{Y}$  takes values in a Borel space  $(R, \mathcal{B})$ , then there is a regular conditional distribution for  $\mathbf{Y}$  given  $\mathcal{D}$ .

*Proof.* By definition, there is a one-to-one mapping  $\varphi: R \leftrightarrow E \in \mathcal{B}_1$  with  $\varphi, \varphi^{-1}$  measurable  $\mathcal{B}, \mathcal{B}_1$  respectively. Take  $\mathbf{Y} = \varphi(\mathbf{Y})$ ;  $\mathbf{Y}$  is a random variable so there is a regular conditional distribution  $\hat{P}_0(A | \mathcal{D}) = P(\mathbf{Y} \in A | \mathcal{D})$  a.s. Define  $\hat{P}(B | \mathcal{D})$ , for  $B \in \mathcal{B}$ , by

$$\hat{P}(B | \mathcal{D}) = \hat{P}_0(\varphi(B) | \mathcal{D}).$$

Because  $\varphi(B)$  is the inverse set mapping of the measurable mapping  $\varphi^{-1}(x)$ ,  $\hat{P}(\cdot | \mathcal{D})$  is a probability on  $\mathcal{B}$  for each  $\omega$ , and is also a version of  $P(\mathbf{Y} \in B | \mathcal{D})$  for every  $B \in \mathcal{B}$ .

Since the distribution of processes is determined on their range space, a regular conditional distribution will suit us just as well as a regular conditional probability. For instance,

**Proposition 4.35.** Let  $\mathbf{Y}$  be a random vector taking values in  $(R, \mathcal{B})$ ,  $\mathbf{y}$  a point in  $R$ ,  $\varphi$  any  $\mathcal{B}$ -measurable function such that  $E|\varphi(\mathbf{Y})| < \infty$ . Then if  $\hat{P}(\cdot | \mathcal{D})$  is a regular conditional distribution for  $\mathbf{Y}$  given  $\mathcal{D}$ ,

$$E(\varphi(\mathbf{Y}) | \mathcal{D}) = \int \varphi(\mathbf{y}) \hat{P}(d\mathbf{y} | \mathcal{D}) \quad \text{a.s.}$$

*Proof.* Same as 4.28 above.

If  $\mathbf{Y}, \mathbf{X}$  are two random vectors taking values in  $(R, \mathcal{B}), (S, \mathcal{E})$  respectively, then define a regular conditional distribution for  $\mathbf{Y}$  given  $\mathbf{X} = \mathbf{x}$ ,  $\hat{P}(B | \mathbf{X} = \mathbf{x})$ , in the analogous way: for each  $\mathbf{x} \in S$ , it is a probability on  $\mathcal{B}$ , and for  $B$  fixed it is a version of  $\hat{P}(B | \mathbf{X} = \mathbf{x})$ . Evidently the results 4.34 and 4.35 hold concerning  $\hat{P}(B | \mathbf{X} = \mathbf{x})$ . Some further useful results are:

**Proposition 4.36.** Let  $\varphi(\mathbf{x}, \mathbf{y})$  be a random variable on the product space  $(R, \mathcal{B}) \times (S, \mathcal{E})$ ,  $E|\varphi(\mathbf{X}, \mathbf{Y})| < \infty$ . If  $\hat{P}(\cdot | \mathbf{X} = \mathbf{x})$  is a regular conditional distribution for  $\mathbf{Y}$  given  $\mathbf{X} = \mathbf{x}$ , then,

$$(4.37) \quad E(\varphi(\mathbf{X}, \mathbf{Y}) | \mathbf{X} = \mathbf{x}) = \int \varphi(\mathbf{x}, \mathbf{y}) \hat{P}(d\mathbf{y} | \mathbf{X} = \mathbf{x}) \quad \text{a.s.}$$

[Note: The point of 4.36 is that  $\mathbf{x}$  is held constant in the integration occurring in the right-hand side of (4.37). Since  $\varphi(\mathbf{x}, \mathbf{y})$  is jointly measurable in  $(\mathbf{x}, \mathbf{y})$ , for fixed  $\mathbf{x}$  it is a measurable function of  $\mathbf{y}$ .]

*Proof.* Let  $\varphi(\mathbf{x}, \mathbf{y}) = \chi_C(\mathbf{x})\chi_D(\mathbf{y})$ ,  $C, D$  measurable sets, then, by Problem 13(b). a.s.  $\hat{P}$ ,

$$\begin{aligned} E(\chi_C(\mathbf{X})\chi_D(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}) &= \chi_C(\mathbf{x})E(\chi_D(\mathbf{Y}) \mid \mathbf{X} = \mathbf{x}) \\ &= \chi_C(\mathbf{x})P(\mathbf{Y} \in D \mid \mathbf{X} = \mathbf{x}). \end{aligned}$$

On the right in 4.37 is

$$\chi_C(\mathbf{x}) \int \chi_D(\mathbf{y}) \hat{P}(d\mathbf{y} \mid \mathbf{X} = \mathbf{x}) = \chi_C(\mathbf{x}) \hat{P}(D \mid \mathbf{X} = \mathbf{x})$$

which, by definition of  $\hat{P}$ , verifies (4.37) for this  $\varphi$ .

Now to finish the proof, just approximate in the usual way; that is, (4.37) is now true for all  $\varphi(\mathbf{x}, \mathbf{y})$  of the form  $\sum_i \chi_{C_i}(\mathbf{x})\chi_{D_i}(\mathbf{y})$ ,  $C_i, D_i$  measurable sets, and apply now the usual monotonicity argument.

One can see here the usefulness of a regular conditional distribution. It is tempting to replace the right-hand side of (4.37) by  $E(\varphi(\mathbf{x}, \mathbf{Y}) \mid \mathbf{X} = \mathbf{x})$ . But this object cannot be defined through the standard definition of conditional expectation (4.18).

A useful corollary is

**Corollary 4.38.** Let  $\mathbf{X}, \mathbf{Y}$  be independent random vectors. For  $\varphi(\mathbf{x}, \mathbf{y})$  as in 4.36,

$$E(\varphi(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X} = \mathbf{x}) = E\varphi(\mathbf{x}, \mathbf{Y}) \quad \text{a.s.}$$

*Proof.* A regular conditional distribution for  $\mathbf{Y}$  given  $\mathbf{X} = \mathbf{x}$  is  $P(\mathbf{Y} \in B)$ . Apply this in 4.36.

## Problems

**16.** Let  $I$  be any interval in  $R^{(1)}$ . A function  $\varphi(x)$  on  $I$  measurable  $\mathcal{B}_1(I)$  is called *convex* if for all  $t \in [0, 1]$  and  $x \in I, y \in I$

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y).$$

Prove that if  $\mathbf{Y}$  is a random variable with range in  $I$ , and  $E|\mathbf{Y}| < \infty$ , then for  $\varphi(x)$  convex on  $I$ , and  $E|\varphi(\mathbf{Y})| < \infty$ ,

- a)  $\varphi(E\mathbf{Y}) \leq E[\varphi(\mathbf{Y})]$  (Jensen's inequality),
- b)  $\varphi(E(\mathbf{Y} \mid \mathcal{D})) \leq E(\varphi(\mathbf{Y}) \mid \mathcal{D})$  a.s.

[On (b) use the existence of a regular conditional distribution.]

17. Let  $\mathbf{X}, \mathbf{Y}$  be independent random vectors.

- a) Show that a regular conditional probability for  $\mathbf{Y}$  given  $\mathbf{X} = \mathbf{x}$  is  $P(\mathbf{Y} \in B)$ .
- b) If  $\varphi(\mathbf{x}, \mathbf{y})$  is a random variable on the product of the range spaces, show that a regular conditional distribution for  $Z = \varphi(\mathbf{X}, \mathbf{Y})$  given  $\mathbf{X} = \mathbf{x}$  is  $P(\varphi(\mathbf{x}, \mathbf{Y}) \in B)$ .

## NOTES

The modern definition of conditional probabilities and expectations, using the Radon-Nikodym theorem, was formulated by Kolmogorov in his monograph of 1933 [98].

The difficulty in getting a regular conditional probability on  $\mathcal{F}(\mathbf{X})$ , given any  $\sigma$ -field  $\mathcal{D}$ , is similar to the extension problem. Once  $\hat{P}(\cdot | \mathcal{D})$  is gotten on the range space  $(R, \mathcal{B})$  of  $\mathbf{X}$ , if for any set  $B \in \mathcal{B}$  contained in the complement of the range of  $\mathbf{X}$ ,  $\hat{P}(B | \mathcal{D}) \equiv 0$ , then by 2.12 we can get a regular conditional probability  $P^*(\cdot | \mathcal{D})$  on  $\mathcal{F}(\mathbf{X})$  given  $\mathcal{D}$ . In this case it is sufficient that the range of  $\mathbf{X}$  be a set in  $\mathcal{B}$ . Blackwell's article [8] also deals with this problem.

The counterexample referred to in the text is this: Let  $\Omega = [0, 1]$ ,  $\mathcal{D} = \mathcal{B}([0, 1])$ . Take  $C$  to be a nonmeasurable set of outer Lebesgue measure one and inner Lebesgue measure zero. The smallest  $\sigma$ -field  $\mathcal{F}$  containing  $\mathcal{D}$  and  $C$  consists of all sets of the form  $A = (C \cap B_1) \cup (C^c \cap B_2)$ , where  $B_1, B_2$  are in  $\mathcal{D}$ . Define  $P$  on  $\mathcal{F}$  by: If  $A$  has the above form,

$$P(A) = \frac{1}{2}I(B_1) + \frac{1}{2}I(B_2).$$

Because the  $B_1, B_2$  in the definition of  $A$  are not unique, it is necessary to check that  $P$  is well defined, as well as a probability. There does not exist any regular conditional probability on  $\mathcal{F}$  given  $\mathcal{D}$ . This example can be found in Doob [39, p. 624].

## CHAPTER 5

### MARTINGALES

#### 1. GAMBLING AND GAMBLING SYSTEMS

Since probability theory started from a desire to win at gambling, it is only sporting to discuss some examples from this area.

**Example 1.** Let  $Z_1, Z_2, \dots$  be independent,  $Z_i = +1, -1$  with probability  $p, 1 - p$ . Suppose that at the  $n$ th toss, we bet the amount  $b$ . Then we receive the amount  $b$  if  $Z_n = 1$ , otherwise we lose the amount  $b$ . A gambling strategy is a rule which tells us how much to bet on the  $(n + 1)$  toss. To be interesting, the rule will depend on the first  $n$  outcomes. In general, then, a gambling strategy is a sequence of functions  $b_n: \{-1, +1\}^{(n)} \rightarrow [0, \infty)$  such that  $b_n(Z_1, \dots, Z_n)$  is the amount we bet on the  $(n + 1)$  game. If we start with initial fortune  $S_0$ , then our fortune after  $n$  plays,  $S_n$ , is a random variable defined by

$$(5.1) \quad S_{n+1} = S_n + Z_{n+1}b_n(Z_1, \dots, Z_n).$$

We may wish to further restrict the  $b_n$  by insisting that  $b_n(Z_1, \dots, Z_n) \leq S_n$ . Define the time of ruin  $n^*$  as the first  $n$  such that  $S_n = 0$ . One question that is certainly interesting is: What is  $P(n^* < \infty)$ ; equivalently what is the probability of eventual ruin?

There is one property of the sequence of fortunes given by (5.1) that I want to focus on.

**Proposition 5.2.** For  $p = \frac{1}{2}$ ,

$$E(S_{n+1} | S_n, S_{n-1}, \dots, S_1) = S_n \quad \text{a.s., } n = 1, 2, \dots$$

For  $p \leq \frac{1}{2}$ ,

$$E(S_{n+1} | S_n, S_{n-1}, \dots, S_1) \leq S_n \quad \text{a.s., } n = 1, 2, \dots$$

*Proof*

$$\begin{aligned} E(S_{n+1} | Z_n, \dots, Z_1) &= E(S_n + Z_{n+1}b_n(Z_1, \dots, Z_n) | Z_n, \dots, Z_1) \\ &= S_n + b_n(Z_1, \dots, Z_n)E(Z_{n+1}). \end{aligned}$$

The last step follows because  $S_n$  is a function of  $Z_1, \dots, Z_n$ . If  $p = \frac{1}{2}$ , then  $EZ_{n+1} = 0$ , otherwise  $EZ_{n+1} < 0$ , but  $b_n$  is nonnegative. Thus, for  $p = \frac{1}{2}$ ,  $E(S_{n+1} | Z_n, \dots, Z_1) = S_n$  a.s. For  $p \leq \frac{1}{2}$ ,  $E(S_{n+1} | Z_n, \dots, Z_1) \leq S_n$  a.s. Note that  $\mathcal{F}(S_1, \dots, S_n) \subset \mathcal{F}(Z_1, \dots, Z_n)$ . Taking conditional

expectations of both sides above with respect to  $S_1, \dots, S_n$  concludes the proof.

**Example 2.** To generalize Example 1, consider any process  $Z_1, Z_2, \dots$ . With any system of gambling on the successive outcomes of the  $Z_1, Z_2, \dots$  the player's fortune after  $n$  plays,  $S_n$  is a function of  $Z_1, \dots, Z_n$ . We assume that corresponding to any gambling system is a sequence of real-valued functions  $\varphi_n$  measurable  $\mathcal{B}_n$  such that the fortune  $S_n$  is given by

$$\varphi_n(Z_1, \dots, Z_n), \text{ and } E|S_n| < \infty.$$

**Definition 5.3.** The sequence of games  $Z_1, Z_2, \dots$ , under the given gambling system is called *fair* if

$$E(S_{n+1} | S_n, \dots, S_1) = S_n \text{ a.s., } n = 1, 2, \dots,$$

unfavorable if

$$E(S_{n+1} | S_n, \dots, S_1) \leq S_n \text{ a.s., } n = 1, 2, \dots$$

The first function of probability in gambling was involved in computation of odds. We wish to address ourselves to more difficult questions such as: Is there any system of gambling in fair or unfavorable games that will yield  $S_n \rightarrow \infty$  a.s.? When can we assert that  $P(n^* < \infty) = 1$  for a large class of gambling systems? This class of problems is a natural gateway to a general study of processes behaving like the sequences  $S_n$ .

## 2. DEFINITIONS OF MARTINGALES AND SUBMARTINGALES

**Definition 5.4.** Given a process  $X_1, X_2, \dots$ . It will be said to form a martingale (MG) if  $E|X_k| < \infty$ ,  $k = 1, 2, \dots$  and

$$(5.5) \quad E(X_{n+1} | X_n, \dots, X_1) = X_n \text{ a.s., } n = 1, 2, \dots$$

It forms a submartingale (SMG) if  $E|X_k| < \infty$ ,  $k = 1, 2, \dots$ , and

$$(5.6) \quad E(X_{n+1} | X_n, \dots, X_1) \geq X_n \text{ a.s., } n = 1, 2, \dots$$

**Example 3. Sums of independent random variables.** Let  $Y_1, Y_2, \dots$  be a sequence of independent random variables such that  $E|Y_k| < \infty$ , all  $k$ ,  $EY_k = 0$ ,  $k = 1, 2, \dots$ . Define  $X_n = Y_1 + \dots + Y_n$ ; then  $E|X_n| \leq \sum_{k=1}^n E|Y_k| < \infty$ . Furthermore,  $\mathcal{F}(X_1, \dots, X_n) = \mathcal{F}(Y_1, \dots, Y_n)$ , so

$$\begin{aligned} E(X_{n+1} | X_n, \dots, X_1) &= E(X_{n+1} | Y_n, \dots, Y_1) \\ &= E(Y_{n+1} | Y_n, \dots, Y_1) + X_n = X_n. \end{aligned}$$

As mentioned in connection with fortunes, if the appropriate inequalities



hold with respect to larger  $\sigma$ -fields, then the same inequalities hold for the process, that is,

**Proposition 5.7.** *Let  $X_1, X_2, \dots$  be a process with  $E|X_k| < \infty, k = 1, 2, \dots$ . Let  $Y_1, \dots$ , be another process such that  $\mathcal{F}(X_1, \dots, X_n) \subset \mathcal{F}(Y_1, \dots, Y_n)$ ,  $n = 1, 2, \dots$ . If*

$$E(X_{n+1} | Y_n, \dots, Y_1) \geq X_n \quad \text{a.s.,} \\ (=)$$

*then  $X_1, X_2, \dots$  is a SMG(MG).*

*Proof.* Follows from

$$E(E(X_{n+1} | Y_n, \dots, Y_1) | X_n, \dots, X_1) = E(X_{n+1} | X_n, \dots, X_1).$$

[See 4.20 (3).]

For any  $m \geq n$ , if  $X_1, \dots$  is a SMG(MG), then

$$E(X_m | X_n, \dots, X_1) \geq X_n \quad \text{a.s.,} \\ (=)$$

because, for example,

$$E(X_{n+2} | X_n, \dots, X_1) = E(E(X_{n+2} | X_{n+1}, X_n, \dots, X_1) | X_n, \dots, X_1) \\ \geq E(X_{n+1} | X_n, \dots, X_1) \geq X_n \quad \text{a.s.}$$

This remark gives us an equivalent way of defining SMG or MG.

**Proposition 5.8.** *Let  $X_1, X_2, \dots$  be a process with  $E|X_k| < \infty$ , all  $k$ . Then it is a SMG(MG) iff for every  $m \geq n$  and  $A \in \mathcal{F}(X_1, \dots, X_n)$ ,*

$$\int_A X_m dP \geq \int_A X_n dP. \\ (=)$$

*Proof.* By definition,  $\int_A X_m dP = \int_A E(X_m | X_n, \dots, X_1) dP$ .

**Problem 1.** Let  $X_1, X_2, \dots$  be a MG or SMG; let  $I$  be an interval containing the range of  $X_n, n = 1, 2, \dots$ , and let  $\varphi(x)$  be convex on  $I$ .

Prove

- If  $E|\varphi(X_n)| < \infty, n = 1, 2, \dots$  and  $X_1, X_2, \dots$  is a MG, then  $X'_n = \varphi(X_n)$  is a SMG.
- If  $E|\varphi(X_n)| < \infty, n = 1, \dots, \varphi(x)$  also nondecreasing on  $I$ , and  $X_1, X_2, \dots$  a SMG, then  $X'_n = \varphi(X_n)$  is a SMG.

### 3. THE OPTIONAL SAMPLING THEOREM

One of the most powerful theorems concerning martingales is built around the idea of transforming a process by optional sampling. Roughly, the idea is this—starting with a given process  $X_1, X_2, \dots$ , decide on the basis of

observing  $X_1, \dots, X_n$  whether or not  $X_n$  will be the first value of the transformed process; then keep observing until on the basis of your observations a second value of the transformed process is chosen, and so forth. More precisely,

**Definition 5.9.** Let  $X_1, X_2, \dots$  be an arbitrary process. A sequence  $m_1, m_2, \dots$  of integer-valued random variables will be called *sampling variables* if

- a)  $1 \leq m_1 \leq m_2 \leq \dots$ ,
- b)  $\{m_k = j\} \in \mathcal{F}(X_1, \dots, X_j)$ .

Then the sequence of random variables  $\check{X}_n$  defined by  $\check{X}_n = X_{m_n}$  is called the *process derived by optional sampling from the original process*.

**Theorem 5.10.** Let  $X_1, X_2, \dots$  be a SMG (MG),  $m_1, m_2, \dots$  sampling variables, and  $\check{X}_n$  the optional sampling process derived from  $X_1, X_2, \dots$ . If

- a)  $E |\check{X}_k| < \infty$ , all  $k$ ,
- b)  $\lim_{N \rightarrow \infty} \int_{\{m_n > N\}} |X_N| dP = 0$ , all  $n$ ,

then the  $\check{X}_1, \check{X}_2, \dots$  process is a SMG(MG).

*Proof.* Let  $A \in \mathcal{F}(\check{X}_1, \dots, \check{X}_n)$ . We must show that

$$\int_A \check{X}_{n+1} dP \geq \int_A \check{X}_n dP.$$

(=)

Let  $D_j = A \cap \{m_n = j\}$ . Then  $A = \cup D_j$  and it suffices to show that

$$\int_{D_j} \check{X}_{n+1} dP \geq \int_{D_j} \check{X}_n dP, \quad \text{all } j.$$

(=)

We assert that  $D_j \in \mathcal{F}(X_1, \dots, X_j)$ , since, letting  $A = \{(\check{X}_1, \dots, \check{X}_n) \in B\}$ ,  $B \in \mathcal{B}_n$ , then

$$\begin{aligned} D_j &= \{(\check{X}_1, \dots, \check{X}_n) \in B, m_n = j\} \\ &= \bigcup_{\substack{j_1, \dots, j_{n-1} \\ j_1 \leq \dots \leq j_{n-1} \leq j}} \{(\check{X}_1, \dots, \check{X}_n) \in B, m_1 = j_1, \dots, m_{n-1} = j_{n-1}, m_n = j\} \\ &= \bigcup_{\substack{j_1, \dots, j_{n-1} \\ j_1 \leq \dots \leq j_{n-1} \leq j}} \{(X_{j_1}, \dots, X_{j_{n-1}}, X_j) \in B, m_1 = j_1, \dots, m_n = j\}. \end{aligned}$$

Evidently, each set in this union is in  $\mathcal{F}(X_1, \dots, X_j)$ . Of course,

$$\int_{D_j} \check{X}_n dP = \int_{D_j} X_j dP.$$

Now, for arbitrary  $N > j$ ,

$$\begin{aligned} \int_{D_j} \check{X}_{n+1} dP &= \sum_{i=j}^N \int_{D_j \cap \{m_{n+1}=i\}} \check{X}_{n+1} dP + \int_{D_j \cap \{m_{n+1} > N\}} \check{X}_{n+1} dP \\ &= \sum_{i=j}^N \int_{D_j \cap \{m_{n+1}=i\}} X_i dP + \int_{D_j \cap \{m_{n+1} > N\}} X_N dP \\ &\quad - \int_{D_j \cap \{m_{n+1} > N\}} (X_N - \check{X}_{n+1}) dP. \end{aligned}$$

The first two terms on the right in the above telescope. Starting with the last term of the sum,

$$\int_{D_j \cap \{m_{n+1}=N\}} X_N dP + \int_{D_j \cap \{m_{n+1} > N\}} X_N dP = \int_{D_j \cap \{m_{n+1} \geq N\}} X_N dP.$$

But  $\{m_{n+1} \geq N\} = \{m_{n+1} < N\}^c \in \mathcal{F}(X_1, \dots, X_{N-1})$ . By 5.8,

$$\int_{D_j \cap \{m_{n+1} \geq N\}} X_N dP \geq \int_{D_j \cap \{m_{n+1} \geq N\}} X_{N-1} dP = \int_{D_j \cap \{m_{n+1} > N-1\}} X_{N-1} dP.$$

Hence the first two terms reduce to being greater than or equal to

$$\int_{D_j \cap \{m_{n+1} \geq j\}} X_j dP.$$

But  $D_j \subset \{m_{n+1} \geq j\}$ , so we conclude that

$$\int_{D_j} \check{X}_{n+1} dP \geq \int_{D_j} \check{X}_n dP - \int_{D_j \cap \{m_{n+1} > N\}} (X_N - \check{X}_{n+1}) dP.$$

Letting  $N \rightarrow \infty$  through an appropriate subsequence, by (b) we can force

$$\int_{D_j \cap \{m_{n+1} > N\}} X_N dP \rightarrow 0.$$

But  $\{m_{n+1} > N\} \downarrow \emptyset$ , and the bounded convergence theorem gives

$$\int_{D_j \cap \{m_{n+1} > N\}} \check{X}_{n+1} dP \rightarrow 0,$$

since  $E|\check{X}_{n+1}| < \infty$  by (a). Q.E.D.

**Corollary 5.11.** Assume the conditions of Theorem 5.10 are in force, and that in addition,  $\overline{\lim} E|X_n| < \infty$ , then

- 1)  $EX_1 \leq E\check{X}_n \leq \overline{\lim} EX_n$ ,
- 2)  $E|\check{X}_n| \leq 2 \overline{\lim} E|X_n| - EX_1$ .

*Proof.* Without loss of generality, take  $m_1 \equiv 1$ . By 5.10,

$$E(\check{X}_n | \check{X}_1) \geq \check{X}_1 \Rightarrow E\check{X}_n \geq E\check{X}_1 = EX_1.$$

Also,

$$E\check{X}_n = \lim_N \sum_{j=1}^N \int_{\{m_n=j\}} X_j dP,$$

but

$$\sum_{j=1}^N \int_{\{m_n=j\}} X_j dP \leq \sum_{j=1}^N \int_{\{m_n=j\}} X_N dP = EX_N - \int_{\{m_n > N\}} X_N dP.$$

By condition (b), we may take a subsequence of  $N$  such that

$$\int_{\{m_n > N\}} X_N dP \rightarrow 0.$$

Using this subsequence, the conclusion  $E\check{X}_n \leq \overline{\lim} EX_n$  follows. For part (2), note that if  $X_n$  is a SMG, so is  $X_n^+ = \max(0, X_n)$ . Thus by 5.10 so is  $\check{X}_n^+$ . Applying (1) we have

$$E\check{X}_n^+ \leq \overline{\lim} EX_n^+.$$

For the original process,

$$E\check{X}_n^+ - E\check{X}_n^- \geq EX_1 \quad \text{or} \quad E\check{X}_n^- \leq E\check{X}_n^+ - EX_1,$$

$$E|\check{X}_n| = E\check{X}_n^+ + E\check{X}_n^- \leq 2 \overline{\lim} EX_n^+ - EX_1 \leq 2 \overline{\lim} E|X_n| - EX_1.$$

**Proposition 5.12.** Under hypothesis (b) of Theorem 5.10, if  $\overline{\lim} E|X_n| < \infty$ , then  $\sup_n E|\check{X}_n| < \infty$ , so that (a) holds and the theorem is in force.

*Proof.* We introduce truncated sampling variables: for  $M > 0$ ,

$$m_{n,M} = \begin{cases} m_n & \text{if } m_n \leq M, \\ M & \text{if } m_n > M, \end{cases}$$

and  $\check{X}_{n,M} = X_{m_{n,M}}$ . The reason for this is

$$E|\check{X}_{n,M}| = \sum_{j=1}^M \int_{\{m_{n,M}=j\}} |X_j| dP \leq \sum_{j=1}^M E|X_j| < \infty,$$

so that the conditions of Theorem 5.10 are in force for  $\check{X}_{n,M}$ . By Corollary 5.11,  $E|\check{X}_{n,M}| \leq \alpha < \infty$ , where  $\alpha$  does not depend on  $n$  or  $M$ . Note now that  $\lim_M m_{n,M} = m_n$ ; hence  $\lim_M \check{X}_{n,M} = \check{X}_n$ . By the Fatou lemma (Appendix A.27),

$$\int \lim_M |\check{X}_{n,M}| dP \leq \lim_M E|\check{X}_{n,M}| \leq \alpha$$

or

$$E|\check{X}_n| \leq \alpha, \quad \text{all } n.$$

One nice application of optional sampling is

**Proposition 5.13.** *If  $X_1, X_2, \dots$  is a SMG then for any  $x > 0$*

- 1) 
$$P\left(\max_{1 \leq j \leq k} X_j > x\right) \leq \frac{1}{x} E|X_k|,$$
- 2) 
$$P\left(\min_{1 \leq j \leq k} X_j < -x\right) \leq \frac{1}{x} (E|X_k| - EX_1).$$

*Proof.* 1) Define sampling variables by

$$m_1 = \begin{cases} \text{first } j \leq k \text{ such that } X_j > x, & \text{or} \\ k & \text{if no such } j \text{ exists.} \end{cases}$$

$$m_n = k, \quad n \geq 2.$$

The conditions of 5.10 are satisfied, since  $\{m_n > N\} = \emptyset$ ,  $N \geq k$ , and  $E|\check{X}_n| \leq \sum_1^k E|X_j| < \infty$ . Now

$$P\left(\max_{1 \leq j \leq k} X_j > x\right) = P(X_{m_1} > x) \leq \frac{1}{x} \int_{\{X_{m_1} > x\}} X_{m_1} dP.$$

By the optional sampling theorem,

$$\int_{\{X_{m_1} > x\}} X_{m_1} dP \leq \int_{\{X_{m_1} > x\}} X_k dP \leq E|X_k|.$$

2) To go below, take  $x > 0$  and define

$$m_1 = 1,$$

$$m_2 = \begin{cases} \text{first } j \leq k \text{ such that } X_j < -x, & \text{or} \\ k & \text{if no such } j \text{ exists,} \end{cases}$$

$$m_n = k, \quad n \geq 3.$$

The conditions of 5.10 are again satisfied, so

$$EX_{m_2} \geq EX_{m_1} = EX_1.$$

But

$$EX_{m_2} = \int_{\{X_{m_2} \geq -x\}} X_{m_2} dP + \int_{\{X_{m_2} < -x\}} X_{m_2} dP$$

$$\leq E|X_k| - xP(X_{m_2} < -x).$$

Therefore, since

$$P\left(\min_{1 \leq j \leq k} X_j < -x\right) = P(X_{m_2} < -x),$$

part (2) follows.

**Problem 2.** Use 5.13 and Problem 1 to prove Kolmogorov's extension of the Chebyshev inequality (see the notes to Chapter 3). That is, if  $X_1, X_2, \dots$  is a MG and  $EX_n^2 < \infty$ , then show

$$P\left(\max_{1 \leq k \leq n} |X_k| > \epsilon\right) \leq \frac{1}{\epsilon^2} EX_n^2.$$

#### 4. THE MARTINGALE CONVERGENCE THEOREM

One of the outstanding strong convergence theorems is

**Theorem 5.14.** Let  $X_1, X_2, \dots$  be a SMG such that  $\overline{\lim} E|X_n| < \infty$ , then there exists a random variable  $X$  such that  $X_n \rightarrow X$  a.s. and  $E|X| < \infty$ .

*Proof.* To prove this, we are going to define sampling variables. Let  $b > a$  be any two real numbers. Let

$$m_1^* = \begin{cases} +\infty & \text{if } X_n > a, \text{ all } n, \\ \text{first } n & \text{such that } X_n \leq a, \text{ otherwise;} \end{cases}$$

$$m_2^* = \begin{cases} +\infty & \text{if } X_n < b, \text{ all } n \geq m_1^*, \\ \text{first } n \geq m_1^* & \text{such that } X_n \geq b, \text{ otherwise.} \end{cases}$$

In general,

$$m_{2n+1}^* = \begin{cases} +\infty & \text{if } X_n > a, \text{ all } n \geq m_{2n}^*, \\ \text{first } n \geq m_{2n}^* & \text{such that } X_n \leq a \text{ otherwise;} \end{cases}$$

$$m_{2n+2}^* = \begin{cases} +\infty & \text{if } X_n < b, \text{ all } n \geq m_{2n+1}^*, \\ \text{first } n \geq m_{2n+1}^* & \text{such that } X_n \geq b, \text{ otherwise.} \end{cases}$$

That is, the  $m_n^*$  are the successive times that  $X_n$  drops below  $a$  or rises above  $b$ . Now, what we would like to conclude for any  $b > a$  is that

$$P(X_n \leq a \text{ i.o., } X_n \geq b \text{ i.o.}) = 0.$$

This is equivalent to proving that if

$$S = \bigcap_1^\infty \{m_n^* < \infty\},$$

that  $P(S) = 0$ . Suppose we defined  $\check{X}_n = X_{m_n^*}$ , and

$$Z = (\check{X}_3 - \check{X}_2) + (\check{X}_5 - \check{X}_4) + \dots$$

On  $S$ ,  $\check{X}_{2n+1} - \check{X}_{2n} \leq -(b-a)$ , so  $Z = -\infty$ , but if  $\check{X}_n$  is a SMG,  $E(\check{X}_{2n+1} - \check{X}_{2n}) \geq 0$ , so  $EZ \geq 0$ , giving a contradiction which would be

resolved only by  $P(S) = 0$ . To make this idea firm, define

$$m_{n,M} = \min(M, m_n^*), \quad \check{X}_{n,M} = X_{m_{n,M}}.$$

Then the optional sampling theorem is in force, and  $\check{X}_{n,M}$  is a SMG. Define

$$Z_M = (\check{X}_{3,M} - \check{X}_{2,M}) + (\check{X}_{5,M} - \check{X}_{4,M}) + \cdots,$$

noting that  $\check{X}_{2n+1,M} - \check{X}_{2n,M} = 0$  for  $m_{2n}^* \geq M$ , which certainly holds for  $2n \geq M$ .

Let  $\beta_M$  be the largest  $n$  such that  $m_{2n}^* < M$ . On the set  $\{\beta_M = k\}$ ,

$$\begin{aligned} Z_M &= (\check{X}_{3,M} - \check{X}_{2,M}) + \cdots + (\check{X}_{2k+1,M} - \check{X}_{2k,M}) \\ &= (\check{X}_{3,M} - \check{X}_{2,M}) + \cdots + (a - \check{X}_{2k,M}) + (\check{X}_{2k+1,M} - a). \end{aligned}$$

The last term above can be positive only if  $m_{2k+1}^* \geq M$ , in which case it becomes  $(X_M - a)$ . In any case, on  $\{\beta_M = k\}$

$$Z_M \leq -k(b - a) + (X_M - a)^+,$$

or, in general,

$$(5.15) \quad Z_M \leq -\beta_M(b - a) + (X_M - a)^+.$$

On the other hand, since  $X_{n,M}$  is a SMG,  $EZ_M \geq 0$ . Taking expectations of (5.15) gives

$$(5.16) \quad E\beta_M \leq \frac{E(X_M - a)^+}{b - a}.$$

Now  $\beta_M$  is nondecreasing, and  $\beta_M \uparrow \infty$  on  $S$ . This contradicts (5.16) since  $\overline{\lim} E(X_M - a)^+ \leq \overline{\lim} E|X_M| + a$ , unless  $P(S) = 0$ . This establishes

$$P(\cup \{\underline{\lim} X_n \leq a < b \leq \overline{\lim} X_n\}) = 0.$$

where the union is taken over all rational  $a, b$ . Then either a random variable  $X$  exists such that  $X_n \rightarrow X$  a.s. or  $|X_n| \rightarrow \infty$  with positive probability. This latter case is eliminated by Fatou's lemma, that is,

$$\int \underline{\lim} |X_n| dP \leq \underline{\lim} \int |X_n| dP.$$

From this we also conclude that

$$\int |X| dP \leq \underline{\lim} \int |X_n| dP < \infty. \quad \text{Q.E.D.}$$

In the body of the above proof, a result is obtained which will be useful in the future.

**Lemma 5.17.** Let  $X_1, \dots, X_M$  be random variables such that  $E|X_n| < \infty$ ,  $n = 1, \dots, M$  and  $E(X_{n+1} | X_n, \dots, X_1) \geq X_n$ ,  $n = 1, \dots, M-1$ . Let  $\beta_M$  be the number of times that the sequence  $X_1, \dots, X_M$  crosses a finite interval  $[a, b]$  from left to right. Then,

$$E\beta_M \leq \frac{E(X_M - a)^+}{b - a}.$$

**Problem 3.** Now use the martingale convergence theorem to prove that for  $X_1, X_2, \dots$  independent random variables,  $EX_k = 0$ ,  $k = 1, 2, \dots$ ,

$$\sum_1^\infty EX_k^2 < \infty \Rightarrow \sum_1^n X_k \xrightarrow{\text{a.s.}}.$$

## 5. FURTHER MARTINGALE THEOREMS

To go further and apply the basic convergence theorem, we need:

**Definition 5.18.** Let  $X_1, X_2, \dots$  be a process. It is said to be uniformly integrable if  $E|X_k| < \infty$ , all  $k$ , and

$$\lim_{x \uparrow \infty} \lim_n \int_{\{|X_n| > x\}} |X_n| dP = 0.$$

**Proposition 5.19.** Let  $X_1, X_2, \dots$  be uniformly integrable, then  $\overline{\lim} E|X_n| < \infty$ . If  $X_n \rightarrow X$  a.s., then  $E|X - X_n| \rightarrow 0$ , and  $EX_n \rightarrow EX$ .

*Proof.* First of all, for any  $x > 0$ ,

$$E|X_n| \leq x + \int_{\{|X_n| > x\}} |X_n| dP.$$

Hence,

$$\overline{\lim} E|X_n| \leq x + \overline{\lim} \int_{\{|X_n| > x\}} |X_n| dP.$$

But the last term must be finite for some value of  $x$  sufficiently large. For the next item, use Fatou's lemma:

$$\int |X| dP = \int \underline{\lim} |X_n| dP \leq \underline{\lim} \int |X_n| dP.$$

Now, by Egoroff's theorem (Halmos [64, p. 88]), for any  $\epsilon > 0$ , we can take  $A \in \mathcal{F}$  such that  $P(A) < \epsilon$  and  $V_n = |X_n - X| \rightarrow 0$  uniformly on  $A^c$ ,

$$\overline{\lim} EV_n = \overline{\lim} \int_A |X_n - X| dP \leq \overline{\lim} \int_A |X_n| dP + \int_A |X| dP.$$



But bounds can be gotten by writing

$$(5.20) \quad \int_A |X_n| dP = \int_{A \cap \{|X_n| \leq x\}} |X_n| dP + \int_{A \cap \{|X_n| > x\}} |X_n| dP \\ \leq x\epsilon + \int_{\{|X_n| > x\}} |X_n| dP.$$

Taking  $\epsilon \downarrow 0$ , we get

$$\overline{\lim} EV_n \leq \overline{\lim}_n \int_{\{|X_n| > x\}} |X_n| dP + \int_{\{|X| > x\}} |X| dP,$$

since (5.20) holds for  $|X|$  as well as  $|X_n|$ . Now letting  $x \uparrow \infty$  gives the result. For the last result, use  $|EX - EX_n| \leq E|X - X_n|$ .

We use this concept of uniform integrability in the proof of:

**Theorem 5.21.** Let  $Z, Y_1, Y_2, \dots$ , be random variables on  $(\Omega, \mathcal{F}, P)$ , such that  $E|Z| < \infty$ . Then

$$E(Z | Y_1, Y_2, \dots, Y_n) \xrightarrow[1]{a.s.} E(Z | Y_1, Y_2, \dots).$$

(Here  $\xrightarrow[1]{a.s.}$  indicates both a.s. and first mean convergence.)

*Proof.* Let  $X_n = E(Z | Y_1, \dots, Y_n)$ . Then  $E|X_n| \leq E|Z|$ . By 4.20(3)

$$E(X_{n+1} | Y_1, \dots, Y_n) = X_n \quad \text{a.s.}$$

Since

$$\mathcal{F}(X_1, \dots, X_n) \subset \mathcal{F}(Y_1, \dots, Y_n),$$

the sequence  $X_1, X_2, \dots$  is a MG. Since  $\overline{\lim} E|X_n| < \infty$ ,

$$X_n \xrightarrow{a.s.} X, \quad \text{and} \quad E|X| \leq \overline{\lim} E|X_n| \leq E|Z|.$$

By convexity,  $|X_n| \leq E(|Z| | Y_1, \dots, Y_n)$ . Hence

$$\int_{\{|X_n| > x\}} |X_n| dP \leq \int_{\{|X_n| > x\}} E(|Z| | Y_1, \dots, Y_n) dP = \int_{\{|X_n| > x\}} |Z| dP \\ \leq \int_{\{\sup_n |X_n| > x\}} |Z| dP.$$

Now,

$$U = \sup_n |X_n|$$

is a.s. finite; hence as  $x \uparrow \infty$ , the sets  $\{U > x\}$  converge down to a set of probability zero. Thus the  $X_n$  sequence is uniformly integrable and  $X_n \xrightarrow{1} X$ . Let  $A \in \mathcal{F}(Y_1, \dots, Y_N)$ ; then

$$\lim_n \int_A X_n dP = \int_A X dP.$$

But for  $n \geq N$ ,

$$\int_A X_n dP = \int_A Z dP.$$

This implies that  $X = Z$  a.s.

**Corollary 5.22.** If  $\mathcal{F}(Z) \subset \mathcal{F}(Y_1, Y_2, \dots)$ , and  $E|Z| < \infty$ , then

$$E(Z | Y_1, \dots, Y_n) \xrightarrow{\text{a.s.}} Z.$$

In particular, if  $A \in \mathcal{F}(Y_1, \dots)$ , then

$$P(A | Y_1, \dots, Y_n) \xrightarrow{\text{a.s.}} \chi_A(\omega).$$

*Proof.* Immediate.

There is a converse of sorts, most useful, to 5.22.

**Theorem 5.23.** Let  $X_1, X_2, \dots$  be a SMG (MG), and uniformly integrable.

Then  $X_n \xrightarrow{\text{a.s.}} X$ , and

$$X_n \leq E(X | X_1, \dots, X_n)$$

(with equality holding if  $X_1, X_2, \dots$  is MG).

*Proof.* By definition, for every  $m > n$ ,

$$E(X_m | X_1, \dots, X_n) \geq X_n \text{ a.s.} \\ (=)$$

Thus, for every set  $A \in \mathcal{F}(X_1, \dots, X_n)$ , and  $m > n$ ,

$$\int_A X_n dP \leq \int_A X_m dP. \\ (=)$$

But by (5.19)  $\overline{\lim} E|X_n| < \infty$ , hence  $X_n \xrightarrow{\text{a.s.}} X$ , implying  $X_n \xrightarrow{1} X$ , so

$$\int_A X_n dP \leq \int_A X dP, \Rightarrow X_n \leq E(X | X_1, \dots, X_n) \text{ a.s.} \\ (=)$$

Therefore, every uniformly integrable MG sequence is a sequence of conditional expectations and every uniformly integrable SMG sequence is bounded above by a MG sequence.

In 5.21 we added conditions, that is, we took the conditional expectation of  $Z$  relative to increasing sequence  $\sigma$ -fields  $\mathcal{F}(Y_1, \dots, Y_n)$ . We can also go the other way.

**Theorem 5.24.** Let  $Z, Y_1, \dots$  be random variables,  $E|Z| < \infty$ ,  $\mathcal{F}$  the tail  $\sigma$ -field on the  $Y_1, Y_2, \dots$  process. Then

$$E(Z | Y_n, Y_{n+1}, \dots) \xrightarrow[1]{\text{a.s.}} E(Z | \mathcal{F}).$$

*Proof.* A bit of a sly dodge is used here. Define  $X_{-n} = E(Z \mid Y_n, Y_{n+1}, \dots)$ ,  $n = 1, 2, \dots$ . Note that

$$E(X_{-n} \mid Y_{n+1}, \dots) = X_{-n-1},$$

so Lemma 5.17 is applicable to the sequence  $X_{-M}, X_{-M+1}, \dots, X_{-1}$ . For any interval  $[a, b]$ , if  $\beta_M$  is the number of times that this sequence crosses from below  $a$  to above  $b$ , then

$$E\beta_M \leq \frac{E(X_{-1} - a)^+}{b - a}.$$

By the same argument as in the main convergence theorem,  $X_{-n}$  either must converge to a random variable  $X$  a.s. or  $|X_{-n}| \rightarrow \infty$  with positive probability. But  $E|X_{-n}| \leq E|Z|$ , so that Fatou lemma again gives  $X_{-n} \rightarrow X$  a.s., and  $E|X| \leq E|Z|$ . Just as in the proof of 5.21,

$$\int_{\{|X_{-n}| > x\}} |X_{-n}| dP \leq \int_{\{|X_{-n}| > x\}} |Z| dP \leq \int_{\{\sup_{n \geq 1} |X_{-n}| > x\}} |Z| dP,$$

so the  $X_{-n}$  sequence is uniformly integrable; hence  $E(Z \mid Y_n, Y_{n+1}, \dots) \xrightarrow{1} X$ . Since  $X$  is an a.s. limit of random variables measurable with respect to  $\mathcal{F}(Y_n, \dots)$ , then  $X$  is measurable with respect to  $\mathcal{F}(Y_n, \dots)$  for every  $n$ . Hence  $X$  is a random variable on  $(\Omega, \mathcal{F})$ . Let  $A \in \mathcal{F}$ ; then  $A \in \mathcal{F}(Y_n, \dots)$ , and

$$\int_A X dP = \lim_n \int_A E(Z \mid Y_n, \dots) dP = \int_A Z dP,$$

which proves that  $X = E(Z \mid \mathcal{F})$  a.s.

## Problems

4. Show that  $E|X_n| < \infty$ ,  $E|X| < \infty$ ,  $E|X_n - X| \rightarrow 0 \Rightarrow X_1, X_2, \dots$  is a uniformly integrable sequence. Get the same conclusion if  $X_n \xrightarrow{\text{a.s.}} X$ , and  $E|X_n| \rightarrow E|X|$ .

5. Apply Theorem 5.21 to the analytic model for coin-tossing and the coin-tossing variables defined in Chapter 1, Section 5. Take  $f(x)$  any Borel measurable function on  $[0, 1)$  such that  $\int |f(x)| dx < \infty$ . Let  $I_1, I_2, \dots, I_N$  be the intervals

$$\left[0, \frac{1}{2^n}\right), \left[\frac{1}{2^n}, \frac{2}{2^n}\right), \dots,$$

and define the step function  $f_n(x)$  by

$$f_n(x) = \frac{1}{\|I_k\|} \int_{I_k} f(y) dy, \quad x \in I_k, \quad k = 1, \dots, N.$$

Then prove that

$$f_n(x) \xrightarrow{\text{a.s.}} f(x).$$

6. Given a process  $X_1, X_2, \dots$ , abbreviate  $\mathcal{F}_n = \mathcal{F}(X_n, X_{n+1}, \dots)$ . Use 5.24 to show that the tail  $\sigma$ -field  $\mathcal{J}$  on the process has the zero-one property ( $C \in \mathcal{J} \Rightarrow P(C) = 0$  or  $1$ ) iff for every  $A \in \mathcal{F}(X)$ ,

$$\limsup_n \sup_{B \in \mathcal{F}_n} |P(A \cap B) - P(A)P(B)| = 0.$$

7. Use Problem 15, Chapter 4, to prove the strong law of large numbers (3.30) from the martingale results.

## 6. STOPPING TIMES

**Definition 5.25.** Given a process  $X_1, X_2, \dots$ , an extended stopping time is an extended integer-valued random variable  $n^*(\omega) \geq 1$  such that  $\{n^* = j\} \in \mathcal{F}(X_1, \dots, X_j)$ . The process  $\check{X}_1, \check{X}_2, \dots$  derived under stopping is defined by

$$\check{X}_n = \begin{cases} X_n, & \text{if } n \leq n^*(\omega), \\ X_{n^*}, & n > n^*(\omega). \end{cases}$$

If we define variables  $m_n(\omega)$  as  $\min(n, n^*(\omega))$ , then obviously the  $m_n(\omega)$  are optional sampling variables, and the  $\check{X}_n$  defined in 5.25 above are given by  $\check{X}_n = X_{m_n}$ . Hence stopping is a special case of optional sampling. Furthermore,

**Proposition 5.26.** Let  $X_1, X_2, \dots$  be a  $\text{SMG}(\text{MG})$ , and  $\check{X}_1, \check{X}_2, \dots$  be derived under stopping from  $X_1, X_2, \dots$ . Then  $\check{X}_1, \check{X}_2, \dots$  is a  $\text{SMG}(\text{MG})$ .

*Proof.* All that is necessary is to show that conditions (a) and (b) of the optional sampling Theorem 5.10 are met. Now

$$\begin{aligned} E|\check{X}_n| &= \int_{\{n \leq n^*\}} |X_n| dP + \int_{\{n > n^*(\omega)\}} |X_{n^*}| dP \\ &\leq E|X_n| + \sum_{j=1}^{n-1} \int_{\{n^*=j\}} |X_j| dP \leq \sum_{j=1}^n E|X_j|. \end{aligned}$$

Noting that  $m_n(\omega) = \min(n, n^*(\omega)) \leq n$ , we know that  $\{m_n > N\} = \emptyset$ , for  $N \geq n$ , and

$$\int_{\{m_n > N\}} |X_N| dP = 0.$$

Not only does stopping appear as a transformation worthy of study, but it is also a useful tool in proving some strong theorems. For any set  $B \in \mathcal{B}_1$ , and a process  $X_1, X_2, \dots$ , define an extended stopping time by

$$n^*(\omega) = \begin{cases} \text{first } n \text{ such that } X_n \in B, \\ \infty & \text{if no such } n \text{ exists.} \end{cases}$$

Then  $\check{X}_1, \check{X}_2, \dots$  is called the process stopped on  $B$ .

**Proposition 5.27.** Let  $X_1, X_2, \dots$  be a SMG,  $B$  the set  $[a, \infty]$ ,  $a > 0$ . If  $E[\sup_n (X_{n+1} - X_n)^+] < \infty$ , then for  $\check{X}_n$  the process stopped on  $B$ ,

$$\lim_n E |\check{X}_n| < \infty.$$

*Proof.* For any  $n$ ,  $\check{X}_n^+ \leq a + U$ , where

$$U = \sup_n (X_{n+1} - X_n)^+.$$

But  $\check{X}_1 = X_1$ . By the fact that  $\check{X}_n$  is a SMG,  $E\check{X}_n \geq EX_1$ , so  $E\check{X}_n^+ \leq E\check{X}_n + EX_1$ . Thus

$$E |\check{X}_n| \leq 2E\check{X}_n^+ - EX_1 \leq 2a + 2EU - EX_1.$$

**Theorem 5.28.** Let  $X_1, X_2, \dots$  be a MG such that  $E(\sup_n |X_{n+1} - X_n|) < \infty$ . If the sets  $A_1, A_2$  are defined by

$$A_1 = \left\{ \omega; \lim_n X_n(\omega) \text{ exists} \right\},$$

$$A_2 = \left\{ \omega; \overline{\lim}_n X_n(\omega) = +\infty, \underline{\lim}_n X_n(\omega) = -\infty \right\},$$

then  $A_1 \cup A_2 = \Omega$  a.s.

*Proof.* Consider the process  $X_1, X_2, \dots$  stopped on  $[K, \infty]$ . By 5.27 and the basic convergence theorem  $\check{X}_n \xrightarrow{\text{a.s.}} X$ . On the set  $F_K = \{\sup_n X_n < K\}$ ,  $X_n = \check{X}_n$ , all  $n$ . Hence on  $F_K$ ,  $\lim_n X_n$  exists and is finite a.s. Thus this limit exists and is finite a.s. on the set  $\bigcup_{K=1}^{\infty} F_K$ , but this set is exactly the set  $\{\overline{\lim}_n X_n < \infty\}$ . By using now the MG sequence  $-X_1, -X_2, \dots$ , conclude that  $\lim_n X_n$  exists and is finite a.s. on the set  $\{\underline{\lim}_n X_n > -\infty\}$ . Hence  $\lim_n X_n$  exists and is finite for almost all  $\omega$  in the set  $\{\overline{\lim}_n X_n < \infty\} \cup \{\underline{\lim}_n X_n > -\infty\}$ , and the theorem is proved.

This theorem is something like a zero-one law. Forgetting about a set of probability zero, according to this theorem we find that for every  $\omega$ , either  $\lim_n X_n(\omega)$  exists finite, or the sequence  $X_n(\omega)$  behaves badly in the sense that  $\underline{\lim}_n X_n(\omega) = -\infty$ ,  $\overline{\lim}_n X_n(\omega) = +\infty$ . There are some interesting and useful applications. An elegant extension of the Borel-Cantelli lemma valid for arbitrary processes comes first.

**Corollary 5.29** (extended Borel-Cantelli lemma). Let  $Y_1, Y_2, \dots$  be any process, and  $A_n \in \mathcal{F}(Y_1, \dots, Y_n)$ . Then almost surely

$$\left\{ \omega; \omega \in A_n \text{ i.o.} \right\} = \left\{ \omega; \sum_1^{\infty} P(A_{n+1} | Y_n, \dots, Y_1) = \infty \right\}.$$

*Proof.* Let

$$X_{n+1} = \sum_{k=1}^n [\chi_{A_{k+1}} - P(A_{k+1} | Y_k, \dots, Y_1)], \quad n \geq 1.$$

Note that

$$\begin{aligned} E(X_{n+1} | Y_n, \dots, Y_1) &= E([\chi_{A_{n+1}} - P(A_{n+1} | Y_n, \dots, Y_1)] | Y_n, \dots, Y_1) + X_n \\ &= X_n. \end{aligned}$$

Obviously, also  $|X_{n+1}| \leq n$ , so  $\{X_n\}$  is a MG sequence. To boot,

$$|X_{n+1} - X_n| \leq 1,$$

so 5.28 is applicable. Now

$$\{\omega; \omega \in A_n \text{ i.o.}\} = \left\{ \omega; \sum_1^\infty \chi_{A_n}(\omega) = \infty \right\}.$$

Let  $D_1 = \{\lim_n X_n \text{ exists finite}\}$ . Then on  $D_1$ ,

$$\sum_1^\infty \chi_{A_n} = \infty \Leftrightarrow \sum_1^\infty P(A_{n+1} | Y_n, \dots, Y_1) = \infty.$$

Let  $D_2 = \{\underline{\lim} X_n = -\infty, \overline{\lim} X_n = +\infty\}$ , then for all  $\omega \in D_2$ ,

$$\sum_1^\infty \chi_{A_n} = \infty \quad \text{and} \quad \sum_1^\infty P(A_{n+1} | Y_n, \dots, Y_1) = \infty.$$

Since  $P(D_1 \cup D_2) = 1$ , the corollary follows.

Some other applications of 5.28 are in the following problems.

### Problems

**8.** A loose end, which is left over from the random signs problem, is that if  $Y_1, Y_2, \dots$  are independent random variables,  $EY_k = 0$ , by the zero-one law either  $X_n = \sum_1^n Y_k$  converges to a finite limit a.s. or diverges a.s. The nature of the divergence can be gotten from 5.28 in an important special case. Show that if  $Y_1, \dots$  are independent,  $|Y_k| \leq \alpha < \infty$ , all  $k$ ,  $EY_k = 0$ ,  $S_n = Y_1 + \dots + Y_n$ , then either

- a)  $P\left(\lim_n S_n \text{ exists}\right) = 1$ , or  
 b)  $P(\underline{\lim} S_n = -\infty, \overline{\lim} S_n = +\infty) = 1$ .

**9.** For any process  $X_1, X_2, \dots$  and sets  $A, B \in \mathcal{B}_1$ , suppose that  $P(X_m \in B \text{ for at least one } m > n | X_n, \dots, X_1) \geq \delta > 0$ , on  $\{X_n \in A\}$ . Then prove that

$$\{X_n \in A \text{ i.o.}\} \subset \{X_n \in B \text{ i.o.}\} \text{ a.s.}$$

[Let  $F_N = \bigcup_N \{X_m \in B\}$ , and use  $P(F_N | X_n \cdots X_1) \rightarrow \chi_{F_N}$ .]

10. Consider a process  $X_1, X_2, \dots$  taking values in  $[0, \infty)$ . Consider  $\{0\}$  an absorbing state in the sense that  $X_n = 0 \Rightarrow X_{n+m} = 0, m \geq 1$ . Let  $D$  be the event that the process is eventually absorbed at zero, that is,

$$D = \{\exists n \text{ such that } X_n = 0\}.$$

If, for every  $x$  there exists a  $\delta > 0$  such that

$$P(D \mid X_n, \dots, X_1) \geq \delta, \text{ if } X_n \leq x, n = 1, 2, \dots,$$

prove that for almost every sample sequence, either  $X_1, X_2, \dots$  is eventually absorbed, or  $X_n \rightarrow \infty$ .

## 7. STOPPING RULES

**Definition 5.30.** If an extended stopping time  $n^*$  is a.s. finite, call it a stopping time or a stopping rule. The stopped variable is  $X_{n^*}$ .

It is clear that if we define  $m_1(\omega) \equiv 1, m_n(\omega) = n^*(\omega), n \geq 2$ , then the  $m_n$  are optional sampling variables. From the optional sampling theorem we get the interesting

**Corollary 5.31.** Let  $n^*$  be a stopping rule. If  $X_1, X_2, \dots$  is a SMG(MG) and if

$$\text{a) } E |X_{n^*}| < \infty,$$

$$\text{b) } \lim_N \int_{\{n^* > N\}} |X_N| dP = 0,$$

then

$$(5.32) \quad EX_{n^*} \geq EX_1.$$

(=)

*Proof.* Obvious.

The interest of 5.31 in gambling is as follows: If a sequence of games is unfavorable under a given gambling system, then the variables  $-S_n$  form a SMG if  $E |S_n| < \infty$ . Suppose we use some stopping rule  $n^*$  which based on the outcome of the first  $j$  games tells us whether to quit or not after the  $j$ th game. Then our terminal fortune is  $S_{n^*}$ . But if (a) and (b) are in force, then

$$ES_{n^*} \leq ES_1,$$

with equality if the game is fair. Thus we cannot increase our expected fortune by using a stopping rule. Also, in the context of gambling, some illumination can be shed on the conditions (a) and (b) of the optional sampling theorem. Condition (b), which is pretty much the stickler, says roughly that the variables  $m_n$  cannot sample too far out in the sequence too fast. For example, if there are constants  $\alpha_n$  such that  $m_n \leq \alpha_n$ , all  $n$  (even if  $\alpha_n \rightarrow \infty$ ), then (b) is automatically satisfied. A counterexample where (b)

is violated is in the honored rule "stop when you are ahead." Let  $Y_1, Y_2, \dots$  be independent random variables,  $Y_i = \pm 1$  with probability  $\frac{1}{2}$  each. Then  $X_n = Y_1 + \dots + Y_n$  is a MG sequence and represents the winnings after  $n$  plays in a coin-tossing game. From Problem 8,  $\lim X_n = +\infty$  a.s., hence we can define a stopping rule by

$$n^* = \{\text{first } n \text{ such that } X_n = 1\},$$

that is, stop as soon as we win one dollar. If (5.32) were in force, then  $EX_{n^*} = EX_1 = 0$ , but  $X_{n^*} \equiv 1$ . Now (a) is satisfied because  $E|X_{n^*}| = 1$ , hence we must conclude that (b) is violated. This we can show directly. Note that  $|X_n|$  is a SMG sequence, and

$$\{X_1 = -1, X_2 \neq 0, \dots, X_{N-1} \neq 0\} \subset \{n^* > N\}.$$

Therefore

$$\begin{aligned} \int_{\{n^* > N\}} |X_N| dP &\geq \int_{\{X_1 = -1, X_2 \neq 0, \dots, X_{N-1} \neq 0\}} |X_N| dP \\ &\geq \int_{\{X_1 = -1, \dots, X_{N-1} \neq 0\}} |X_{N-1}| dP \\ &= \int_{\{X_1 = -1, \dots, X_{N-2} \neq 0\}} |X_{N-1}| dP. \end{aligned}$$

Going down the ladder we find that

$$\int_{\{n^* > N\}} |X_N| dP \geq \int_{\{X_1 = -1\}} |X_1| dP = \frac{1}{2}.$$

Here, as a matter of fact,  $n^*$  can be quite large. For example, note that  $En^* = \infty$ , because if  $Y_1 = -1$ , we have to wait for an equalization, in other words, wait for the first time that  $Y_2 + \dots + Y_n = 0$  before we can possibly get a dollar ahead. Actually, on the converse side, it is not difficult to prove.

**Proposition 5.33.** *Let  $X_1, X_2, \dots$  be a SMG(MG) and  $n^*$  a stopping rule. If  $En^* < \infty$  and  $E(|X_{n+1} - X_n| | X_n, \dots, X_1) \leq \alpha < \infty$ ,  $n \leq n^*$ , then*

$$EX_{n^*} \geq EX_1.$$

(=)

*Proof.* All we need to do is verify the conditions of 5.31. Denote  $Z_n = |X_n - X_{n-1}|$ ,  $n > 1$ ,  $Z_1 = |X_1|$ ,  $Y = Z_1 + \dots + Z_{n^*}$ . Hence  $|X_{n^*}| \leq Y$ .

$$EY = \sum_{k=1}^{\infty} \int_{\{n^*=k\}} Y dP = \sum_{k=1}^{\infty} \sum_{j=1}^k \int_{\{n^*=k\}} Z_j dP.$$



Interchange the order of summation so that

$$EY = \sum_{j=1}^{\infty} \int_{\{n^* \geq j\}} Z_j dP.$$

The set  $\{n^* \geq j\} = \{n^* < j\}^c$  is in  $\mathcal{F}(X_1, \dots, X_{j-1})$ . Therefore

$$\int_{\{n^* \geq j\}} Z_j dP = \int_{\{n^* \geq j\}} E(Z_j | X_{j-1}, \dots, X_1) dP \leq \alpha P(n^* \geq j),$$

and we get

$$EY \leq \alpha \sum_{j=1}^{\infty} P(n^* \geq j) = \alpha E n^*.$$

For 5.31(b), since  $Z_1 + \dots + Z_N \leq Y$  on the set  $\{n^* > N\}$ ,

$$\int_{\{n^* > N\}} |X_N| dP \leq \int_{\{n^* > N\}} Y dP.$$

As  $N \rightarrow \infty$ ,  $\{n^* > N\} \downarrow \emptyset$  a.s. Apply the bounded convergence theorem to  $\chi_{\{n^* > N\}} Y$  to get the result.

Proposition 5.33 has interesting applications to sums of independent random variables. Let  $Y_1, Y_2, \dots$  be independent and identically distributed,  $S_n = Y_1 + \dots + Y_n$ , and assume that for some real  $\lambda \neq 0$ ,  $\varphi(\lambda) = Ee^{\lambda Y_1}$  exists, and that  $\varphi(\lambda) \geq 1$ . Then

**Proposition 5.34 (Wald's identity).** *If  $n^*$  is a stopping time for the sums  $S_1, S_2, \dots$  such that  $|S_n| < \gamma$ ,  $n \leq n^*$ , and  $E n^* < \infty$ , then*

$$E\left(\frac{e^{\lambda S_{n^*}}}{\varphi(\lambda)^{n^*}}\right) = 1.$$

*Proof.* The random variables

$$X_n = \frac{e^{\lambda S_n}}{\varphi(\lambda)^n}$$

form a MG, since

$$E\left(\frac{e^{\lambda S_{n+1}}}{\varphi(\lambda)^{n+1}} \middle| S_n, \dots, S_1\right) = \frac{e^{\lambda S_n}}{\varphi(\lambda)^{n+1}} E(e^{\lambda Y_{n+1}}) = \frac{e^{\lambda S_n}}{\varphi(\lambda)^n}.$$

Obviously,  $EX_1 = 1$ , so if the second condition of 5.33 holds, then Wald's identity follows. The condition is

$$E\left(\left|\frac{e^{\lambda S_{n+1}}}{\varphi(\lambda)} - e^{\lambda S_n}\right| \middle| S_n, \dots, S_1\right) \leq \alpha \varphi(\lambda)^n, \quad n \leq n^*,$$

or

$$e^{\lambda S_n} E\left|\frac{e^{\lambda Y_1}}{\varphi(\lambda)} - 1\right| \leq \alpha \varphi(\lambda)^n, \quad n \leq n^*$$

which is clearly satisfied under 5.34.

### Problems

11. For  $n^*$  a stopping time for sums  $S_1, S_2, \dots$  of independent, identically distributed random variables  $Y_1, Y_2, \dots$ ,  $E|Y_1| < \infty$ , prove that  $En^* < \infty$  implies that

$$ES_{n^*} = EY_1 \cdot En^*.$$

Use this to give another derivation of Blackwell's equation (Problem 21, Chapter 3).

12. Let  $Y_1, Y_2, \dots$  be independent and equal  $\pm 1$  with probability  $\frac{1}{2}$ . Let  $r, s$  be positive integers. Define

$$n^* = \{\text{first } n \text{ such that } S_n = r \text{ or } -s\}.$$

Show

$$\text{a) } En^* < \infty,$$

$$\text{b) } P(S_{n^*} = r) = \frac{s}{r+s}.$$

For  $r = s$ , evaluate  $Ee^{-\lambda n^*}$ ,  $\lambda > 0$ .

13. (See Doob, [39, p. 308].) For sums of independent, identically distributed random variables  $Y_1, Y_2, \dots$ , define  $n^*$  as the time until the first positive sum, that is,

$$n^* = \min \{n; S_n > 0\}.$$

Prove that if  $EY_1 = 0$ , then  $En^* = \infty$ .

## 8. BACK TO GAMBLING

The reason that the strategy "quit when you are ahead" works in the fair coin-tossing game is that an infinite initial fortune is assumed. That is, there is no lower limit, say  $-M$ , such that if  $S_n$  becomes less than  $-M$  play is ended.

A more realistic model of gambling would consider the sequence of fortunes  $S_0, S_1, S_2, \dots$  (that is, money in hand) as being nonnegative and finite. We now turn to an analysis of a sequence of fortunes under gambling satisfying

(5.35) i)  $S_n \geq 0$ ,  $n = 0, 1, 2, \dots$ ,  $S_0$  constant,

ii)  $ES_n < \infty$  and  $E(S_{n+1} | S_n, \dots, S_0) \leq S_n$  a.s.,  $n \geq 0$ .

In addition, in any reasonable gambling house, and by the structure of our monetary system, if we bet on the  $n$ th trial, there must be a lower bound to the amount we can win or lose. We formalize this by

**Assumption 5.36.** There exists a  $\delta > 0$  such that either

$$S_{n+1} = S_n \quad \text{or} \quad |S_{n+1} - S_n| \geq \delta.$$

**Definition 5.37.** We will say that we bet on the  $n$ th game if  $|S_n - S_{n-1}| \geq \delta$ . Let  $n^*$  be the (possibly extended) time of the last bet, that is,

$$n^* = \begin{cases} \text{largest } n \text{ such that } |S_n - S_{n-1}| \geq \delta, \\ +\infty \text{ if no such } n \text{ exists.} \end{cases}$$

where  $S_0$  is the starting fortune.

Under (5.35), (i) and (ii), and 5.36, the martingale convergence theorem yields strong results. You can't win!

**Theorem 5.38.**

$$P(n^* < \infty) = 1, \quad ES_{n^*} \leq S_0.$$

**Remark.** The interesting thing about  $P(n^* < \infty) = 1$  is the implication that in an unfavorable (or fair) sequence of games, one cannot keep betting indefinitely. There must be a last bet. Furthermore,  $ES_{n^*} \leq S_0$  implies that the expected fortune after the last bet is smaller than the initial fortune.

*Proof.* Let  $X_n = -S_n$ ; then the  $X_0, X_1, \dots$  sequence is a SMG. Furthermore,  $E|X_n| = E|S_n| = ES_n$ . Thus  $E|X_n| \leq S_0$ , all  $n$ . Hence there exists a random variable  $X$  such that  $X_n \xrightarrow{\text{a.s.}} X$ . Thus

$$P(|X_{n+1} - X_n| \geq \delta \text{ i.o.}) = 0,$$

or

$$P(|S_{n+1} - S_n| \geq \delta \text{ i.o.}) = 0 \quad \text{or} \quad P(n^* < \infty) = 1.$$

To prove the second part use the monotone convergence theorem

$$\int S_{n^*} dP = \lim_n \int_{\{n^* \leq n\}} S_{n^*} dP.$$

But on  $\{n^* \leq n\}$ ,  $S_{n^*} = S_n$ ; hence

$$ES_{n^*} = \lim_n \int_{\{n^* \leq n\}} S_n dP \leq \overline{\lim} ES_n \leq S_0.$$

Note that the theorem is actually a simple corollary of the martingale convergence theorem. Now suppose that the gambling house has a minimum bet of  $\alpha$  dollars and we insist on betting as long as  $S_n \geq \alpha$ ; then  $n^*$  becomes the time "of going broke," that is,  $n^* = \{\text{first } n \text{ such that } S_n < \alpha\}$ , and the obvious corollary of 5.38 is that the *persistent gambler goes broke with probability one*.

## NOTES

Martingales were first fully explored by Doob, in 1940 [32] and systematically developed in his book [39] of 1953. Their widespread use in probability theory has mostly occurred since that time. However, many of the results

had been scattered around for some time. In particular, some of them are due to Lévy, appearing in his 1937 book [103], and some to Ville [137, 1939]. Some of the convergence theorems in a measure-theoretic framework are due to Andersen and Jessen. See the Appendix to Doob's book [39] for a discussion of the connection with the Andersen-Jessen approach, and complete references. The important concepts of optional sampling, optional stopping, and the key lemma 5.17 are due to Doob.

David Freedman has pointed out to me that many of the convergence results can be gotten from the inequality 5.13. For example, here is a more elementary and illuminating proof of 5.21 for  $Z$  an  $\mathcal{F}(X)$  measurable random variable. For any  $\epsilon > 0$ , take  $k, Z_k$  measurable  $\mathcal{F}(X_1, \dots, X_k)$  such that

$$E|Z - Z_k| \leq \epsilon.$$

Now,

$$E(Z_k | X_1, \dots, X_n) = Z_k, \quad n \geq k.$$

Let

$$Y_n = E(Z - Z_k | X_1, \dots, X_n),$$

then the  $\{Y_n\}$  is a MG,  $\{|Y_n|\}$  is a SMG, and by 5.13,

$$P\left(\sup_{1 \leq n \leq N} |Y_n| \geq x\right) \leq \frac{1}{x} E|Y_N| \leq \frac{1}{x} E|Z - Z_k|.$$

Thus,

$$P\left(\overline{\lim}_n |E(Z | X_1, \dots, X_n) - Z_k| \geq x\right) \leq \epsilon/x.$$

Take  $\epsilon \downarrow 0$  fast enough so that  $Z_k \xrightarrow{\text{a.s.}} Z$  to get the result

$$E(Z | X_1, \dots, X_n) \xrightarrow{\text{a.s.}} Z.$$

For a fascinating modern approach to gambling strategies, see the book by Dubins and Savage [40].

## CHAPTER 6

# STATIONARY PROCESSES AND THE ERGODIC THEOREM

### 1. INTRODUCTION AND DEFINITIONS

The question here is: Given a process  $X_1, X_2, \dots$ , find conditions for the almost sure convergence of  $(X_1 + \dots + X_n)/n$ . Certainly, if the  $\{X_n\}$  are independent identically distributed random variables and  $E|X_1| < \infty$ , then,

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{a.s.}} EX_1.$$

A remarkable weakening of this result was proved by Birkhoff in 1931 [4]. Instead of having independent identically distributed random variables, think of requiring that the distribution of the process not depend on the placement of the time origin. In other words, assume that no matter when you start observing the sequence of random variables the resulting observations will have the same probabilistic structure.

**Definition 6.1.** A process  $X_1, X_2, \dots$  is called stationary if for every  $k$ , the process  $X_{k+1}, X_{k+2}, \dots$  has the same distribution as  $X_1, X_2, \dots$ , that is, for every  $B \in \mathcal{B}_\infty$ .

$$(6.2) \quad P((X_1, X_2, \dots) \in B) = P((X_{k+1}, X_{k+2}, \dots) \in B).$$

Since the distribution is determined by the distribution functions, (6.2) is equivalent to: For every  $x_1, \dots, x_n$ , and integer  $k > 0$ ,

$$(6.3) \quad P(X_1 < x_1, \dots, X_n < x_n) = P(X_{k+1} < x_1, \dots, X_{k+n} < x_n).$$

In particular, if a process is stationary, then all the one-dimensional distribution functions are the same, that is,

$$P(X_1 < x) = P(X_k < x), \quad k = 1, 2, \dots$$

We can reduce (6.2) and (6.3) by noting

**Proposition 6.4.** A process  $X_1, X_2, \dots$  is stationary if the process  $X_2, X_3, \dots$  has the same distribution as  $X_1, X_2, \dots$ .

*Proof.* Let  $X'_k = X_{k+1}$ ,  $k = 1, 2, \dots$ . Then  $X'_1, X'_2, \dots$  has the same distribution as  $X_1, X_2, \dots$ . Hence  $X'_2, X'_3, \dots$  has the same distribution as  $X'_1, X'_2, \dots$ , and so forth.

Sometimes it is more convenient to look at stationary processes that consist of a double-ended sequence of random variables  $\dots, X_{-1}, X_0, X_1, \dots$ . In this context, what we have is an infinite sequence of readings, beginning in the infinitely remote past and continuing into the infinite future. Define such a process to be stationary if its distribution does not depend on choice of an origin, i.e., in terms of finite dimensional distributions:

$$P(X_1 < x_1, \dots, X_n < x_n) = P(X_{k+1} < x_1, \dots, X_{k+n} < x_n)$$

for all  $x_1, \dots, x_n$  and all  $k$ , both positive and negative.

The interesting point here is

**Proposition 6.5.** *Given any single-ended stationary process  $X_1, X_2, \dots$ , there is a double-ended stationary process  $\dots, \tilde{X}_{-1}, \tilde{X}_0, \tilde{X}_1, \dots$  such that  $\tilde{X}_1, \tilde{X}_2, \dots$  and  $X_1, X_2, \dots$  have the same distribution.*

*Proof.* From the Extension Theorem 2.26, all we need to define the  $\tilde{X}_k$  process is a set of consistent distribution functions; i.e., we need to define

$$P(\tilde{X}_{-m} < x_{-m}, \dots, \tilde{X}_0 < x_0, \dots, \tilde{X}_n < x_n)$$

such that if either  $x_{-m}$  or  $x_n \uparrow \infty$ , then we drop down to the next highest distribution function. We do this by defining

$$P(\tilde{X}_{-m} < x_{-m}, \dots, \tilde{X}_n < x_n) = P(X_1 < x_{-m}, \dots, X_{n+m+1} < x_n),$$

that is, we slide the distribution functions of the  $X_1, \dots$  process to the left. Now  $X_1, X_2, \dots$  can be looked at as the continuation of a process that has already been going on an infinite length of time.

Starting with any stationary process, an infinity of stationary processes can be produced.

**Proposition 6.6.** *Let  $X_1, X_2, \dots$  be stationary,  $\varphi(\mathbf{x})$  measurable  $\mathcal{B}_\infty$ , then the process  $Y_1, Y_2, \dots$  defined by*

$$Y_k = \varphi(X_k, X_{k+1}, \dots)$$

*is stationary.*

*Proof.* On  $R^{(\infty)}$  define  $\varphi_k(\mathbf{x})$  as  $\varphi(x_k, x_{k+1}, \dots)$ . The set

$$A = \{\mathbf{x}; (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \dots) \in B\},$$

$B \in \mathcal{B}_\infty$ , is in  $\mathcal{B}_\infty$ , because each  $\varphi_k(\mathbf{x})$  is a random variable on  $(R^{(\infty)}, \mathcal{B}_\infty)$ . Note

$$\{\omega; (Y_1, Y_2, \dots) \in B\} = \{\omega; (X_1, X_2, \dots) \in A\}$$

and

$$\{\omega; (Y_2, Y_3, \dots) \in B\} = \{\omega; (X_2, X_3, \dots) \in A\},$$

which implies the stationarity of the  $Y_k$  sequence.

**Corollary 6.7.** Let  $X_1, X_2, \dots$  be independent and identically distributed random variables,  $\varphi(\mathbf{x})$  measurable  $\mathcal{B}_\infty$ ; then

$$Y_k = \varphi(X_k, X_{k+1}, \dots)$$

is stationary.

*Proof.* The  $X_1, X_2, \dots$  sequence is stationary.

**Problem 1.** Look at the unit circle and define

$$f(x) = \begin{cases} 1, & 0 \leq x < \pi, \\ 0, & \pi \leq x < 2\pi. \end{cases}$$

Here  $\Omega$  is the unit circle,  $\mathcal{F}$  the Borel  $\sigma$ -field. Take  $P$  to be Lebesgue measure divided by  $2\pi$ . Take  $\theta$  to be an irrational angle. Define  $x_1 = x$ ,  $x_{k+1} = (x_k + \theta)[2\pi]$ , and  $X_k(x) = f(x_k)$ . Hence the process is a sequence of zeros and ones, depending on whether  $x_k$  is in the last two quadrants or first two when  $x_1$  is picked at random on the circumference. Prove that  $X_1, X_2, \dots$  is stationary, ( $[\alpha]$  denotes modulo  $\alpha$ ).

## 2. MEASURE-PRESERVING TRANSFORMATIONS

Consider a probability space  $(\Omega, \mathcal{F}, P)$  and a transformation  $T$  of  $\Omega$  into itself. As usual, we will call  $T$  measurable if the inverse images under  $T$  of sets in  $\mathcal{F}$  are again in  $\mathcal{F}$ ; that is, if  $T^{-1}A = \{\omega; T\omega \in A\} \in \mathcal{F}$ , all  $A \in \mathcal{F}$ .

**Definition 6.8.** A measurable transformation  $T$  on  $\Omega \rightarrow \Omega$  will be called measure-preserving if  $P(T^{-1}A) = P(A)$ , all  $A \in \mathcal{F}$ .

To check whether a given transformation  $T$  is measurable, we can easily generalize 2.28 and conclude that if  $T^{-1}C \in \mathcal{F}$ , for  $C \in \mathcal{C}$ ,  $\mathcal{F}(\mathcal{C}) = \mathcal{F}$ , then  $T$  is measurable. Again, to check measure-preserving, both  $P(T^{-1}A)$  and  $P(A)$  are  $\sigma$ -additive probabilities, so we need check only their agreement on a class  $\mathcal{C}$ , closed under intersections, such that  $\mathcal{F}(\mathcal{C}) = \mathcal{F}$ .

Starting from measure-preserving transformations (henceforth assumed measurable) a large number of stationary processes can be generated. Let  $X(\omega)$  be any random variable on  $(\Omega, \mathcal{F}, P)$ . Let  $T$  be measure-preserving and define a process  $X_1, X_2, \dots$  by  $X_1(\omega) = X(\omega)$ ,  $X_2(\omega) = X(T\omega)$ ,  $X_3(\omega) = X(T^2\omega)$ ,  $\dots$ . Another way of looking at this is: If  $X_1(\omega)$  is some measurement on the system at time one, then  $X_n(\omega)$  is the same measurement after the system has evolved  $n - 1$  steps so that  $\omega \rightarrow T^{n-1}\omega$ . It should be intuitively clear that the distribution of the  $X_1, X_2, \dots$  sequence does not depend on origin, since starting from any  $X_n(\omega)$  we get  $X_{n+1}(\omega)$  as  $X_n(T\omega)$ , and so forth. To make this firm, denote by  $T^0$  the identity operator, and we prove

**Proposition 6.9.** *Let  $T$  be measure preserving on  $(\Omega, \mathcal{F}, P)$ ,  $X$  a random variable on  $(\Omega, \mathcal{F})$ ; then the sequence  $X_n(\omega) = X(T^{n-1}\omega)$ ,  $n = 1, 2, \dots$  is a stationary sequence of random variables.*

*Proof.* First of all,  $X_n(\omega)$  is a random variable, because  $\{X_n(\omega) \in B\} = \{X(T^{n-1}\omega) \in B\}$ . Let  $A = \{X \in B\}$ . Then

$$\{X(T^{n-1}\omega) \in B\} = \{\omega; T^{n-1}\omega \in A\}.$$

Evidently, however,  $T$  measurable implies  $T^{n-1}$  measurable, or  $T^{n-1}A \in \mathcal{F}$ . Now let  $A = \{\omega; (X_1, X_2, \dots) \in B\}$ ,  $B \in \mathcal{B}_\infty$ , thus

$$A = \{\omega; (X(\omega), X(T\omega), \dots) \in B\}.$$

Look at  $A_1 = \{\omega; (X_2, X_3, \dots) \in B\}$ . This similarly is the set  $\{\omega; (X(T\omega), X(T^2\omega), \dots) \in B\}$ . Hence  $\omega \in A_1 \Leftrightarrow T\omega \in A$  or  $A_1 = T^{-1}A$ . But, by hypothesis,  $P(T^{-1}A) = P(A)$ .

Can every stationary process be generated by a measure-preserving transformation? Almost! In terms of distribution, the answer is Yes. Starting from any stationary process  $X_1(\omega), X_2(\omega), \dots$  go to the coordinate representation process  $\hat{X}_1, \hat{X}_2, \dots$  on  $(R^{(\infty)}, \mathcal{B}_\infty, \hat{P})$ . By definition,

$$\hat{X}_n(\mathbf{x}) = x_n.$$

**Definition 6.10.** *On  $(R^{(\infty)}, \mathcal{B}_\infty)$  define the shift transformation  $S: R^{(\infty)} \rightarrow R^{(\infty)}$  by  $S(x_1, x_2, \dots) = (x_2, x_3, \dots)$ .*

So, for example,  $S(3, 2, 7, 1, \dots) = (2, 7, 1, \dots)$ .

The point is that from the definitions,  $X_n(\mathbf{x}) = X_1(S^{n-1}\mathbf{x})$ . We prove below that  $S$  is measurable and measure-preserving, thus justifying the answer of "Almost" above.

**Proposition 6.11.** *The transformation  $S$  defined above is measurable, and if  $X_1, X_2, \dots$  is stationary, then  $S$  preserves  $\hat{P}$  measure.*

*Proof.* To show  $S$  measurable, consider

$$B = S^{-1}\{\mathbf{x}; x_1 \in I_1, \dots, x_n \in I_n\} = S^{-1}C.$$

By definition, letting  $(S\mathbf{x})_k$  be the  $k$ th coordinate of  $S\mathbf{x}$ , we find that

$$B = \{\mathbf{x}; (S\mathbf{x})_1 \in I_1, \dots, (S\mathbf{x})_n \in I_n\} = \{\mathbf{x}; x_2 \in I_1, \dots, x_{n+1} \in I_n\},$$

and that  $B$  is therefore obviously in  $\mathcal{B}_\infty$ . Furthermore, by the stationarity of  $\hat{X}_1, \hat{X}_2, \dots$

$$\begin{aligned} \hat{P}(B) &= \hat{P}(S^{-1}C) = \hat{P}(\hat{X}_2 \in I_1, \dots, \hat{X}_{n+1} \in I_n) \\ &= \hat{P}(\hat{X}_1 \in I_1, \dots, \hat{X}_n \in I_n) = \hat{P}(C), \end{aligned}$$

So  $S$  is also measure-preserving.



### Problems

2. Show that the following transformations are *measurable* and *measure-preserving*.

1)  $\Omega = [0, 1)$ ,  $\mathcal{F} = \mathcal{B}[0, 1)$ ,  $P = dx$ . Let  $\lambda$  be any number in  $[0, 1)$  and define  $Tx = (x + \lambda)[1]$ .

2)  $\Omega = [0, 1)$ ,  $\mathcal{F} = \mathcal{B}[0, 1)$ ,  $P = dx$ ,  $Tx = (2x)[1]$ .

3) Same as (2) above, but  $Tx = (kx)[1]$ , where  $k > 2$  is integral.

3. Show that for the following transformations on  $[0, 1)$ ,  $\mathcal{B}[0, 1)$ , there is no  $P$  such that  $P(\text{single point}) = 0$ , and the transformation preserves  $P$ .

1)  $Tx = \lambda x$ ,  $0 < \lambda < 1$ .

2)  $Tx = x^2$ .

4. On  $\Omega = [0, 1)$ , define  $T: \Omega \rightarrow \Omega$  by  $Tx = (2x)[1]$ . Use  $\mathcal{F} = \mathcal{B}([0, 1))$ ,  $P = dx$ . Define

$$X(x) = \begin{cases} 0, & 0 \leq x < \frac{1}{2}, \\ 1, & \frac{1}{2} \leq x < 1. \end{cases}$$

Show that the sequence  $X_n(x) = X(T^{n-1}x)$  consists of independent zeros and ones with probability  $\frac{1}{2}$  each.

Show that corresponding to every stationary sequence  $X_1(\omega), X_2(\omega), \dots$  such that  $X_n(\omega) \in \{0, 1\}$ , there is a probability  $Q(dx)$  on  $\mathcal{B}[0, 1)$  such that  $Tx = (2x)[1]$  preserves  $Q$ -measure, and such that the  $X_n(x)$  sequence defined above has the same distribution with respect to  $\mathcal{B}[0, 1)$ ,  $Q(dx)$  as  $X_1, X_2, \dots$

### 3. INVARIANT SETS AND ERGODICITY

Let  $T$  be a measure-preserving transformation on  $(\Omega, \mathcal{F}, P)$ .

**Definition 6.12.** A set  $A \in \mathcal{F}$  is *invariant* if  $T^{-1}A = A$ .

If  $A$  is an invariant set, then the motion  $T$  of  $\Omega \rightarrow \Omega$  carries  $A$  into  $A$ ; that is, if  $\omega \in A$ , then  $T\omega \in A$  (because  $T^{-1}A^c = A^c$ ).  $A^c$  is also invariant, and for all  $n$ ,  $T^n$  carries points of  $A$  into  $A$  and points of  $A^c$  into  $A^c$ . Because of the properties of inverse mappings we have

**Proposition 6.13.** The class of invariant sets is a  $\sigma$ -field  $\mathfrak{I}$ .

*Proof.* Just write down definitions.

In the study of dynamical systems,  $\Omega$  is the phase space of the system, and if  $\omega$  is the state of the system at  $t = 0$ , then its state at time  $t$  is given by  $T_t\omega$ , where  $T_t: \Omega \rightarrow \Omega$  is the motion of the phase space into itself induced by the equations of motion. For a conservative system  $T_t(T_s\omega) = T_{t+s}\omega$ .

We discretize time and take  $T = T_1$ , so the state of the system at time  $n$  is given by  $T^n\omega$ . Suppose that  $X(\omega)$  is some observable function of the state  $\omega$ . Physically, in taking measurements, the observation time is quite long compared to some natural time scale of molecular interactions. So we measure, not  $X(\omega)$ , but the average of  $X(\omega)$  over the different states into which  $\omega$  passes with the evolution of the system. That is, we measure

$$\frac{1}{\tau} \int_0^\tau X(T_t\omega) dt,$$

for  $\tau$  large, or in discrete time

$$\frac{1}{n} \sum_{k=1}^n X(T^{k-1}\omega),$$

for  $n$  large. The brilliant insight of Gibbs was the following argument: that in time, the point  $\omega_t = T_t\omega$  wandered all over the phase space and that the density of the points  $\omega_t$  in any neighborhood tended toward a limiting distribution. Intuitively, this limiting distribution of points had to be invariant under  $T$ . If there is such a limiting distribution, say a measure  $P$ , then we should be able to replace the limiting time average

$$\frac{1}{n} \sum_{k=1}^n X(T^{k-1}\omega)$$

by the phase average

$$\int X(\omega)P(d\omega).$$

Birkhoff's result was that this argument, *properly formulated*, was true! To put Gibb's conjecture in a natural setting, take  $\Omega$  to be all points on a surface of constant energy. (This will be a subset of  $R^{6n}$  where  $n$  is the number of particles.) Take  $\mathcal{F}$  to be the intersection of  $\mathcal{B}_{\mathfrak{g}_n}$  with  $\Omega$  and  $P$  the normalized surface area on  $\Omega$ . By Liouville's theorem,  $T: \Omega \rightarrow \Omega$  preserves  $P$ -measure. The point is now that  $T^n\omega$  will never become distributed over  $\Omega$  in accordance with  $P$  if there are invariant subsets  $A$  of  $\Omega$  such that  $P(A) > 0$ ,  $P(A^c) > 0$ , because in this case the points of  $A$  will remain in  $A$ ; similarly, for  $A^c$ . Thus, the only hope for Gibb's conjecture is that every invariant set  $A$  has probability zero or one.

To properly formulate it, begin with

**Definition 6.14.** Let  $T$  be measure-preserving on  $(\Omega, \mathcal{F}, P)$ .  $T$  is called *ergodic* if for every  $A \in \mathfrak{I}$ ,  $P(A) = 0$  or  $1$ .

One question that is relevant here is: Suppose one defined events  $A$  to be *a.s. invariant* if  $P(A \triangle T^{-1}A) = 0$ . Is the class of a.s. invariant events considerably different from the class of invariant events? Not so! It is exactly the completion of  $\mathfrak{I}$  with respect to  $P$  and  $\mathcal{F}$ .

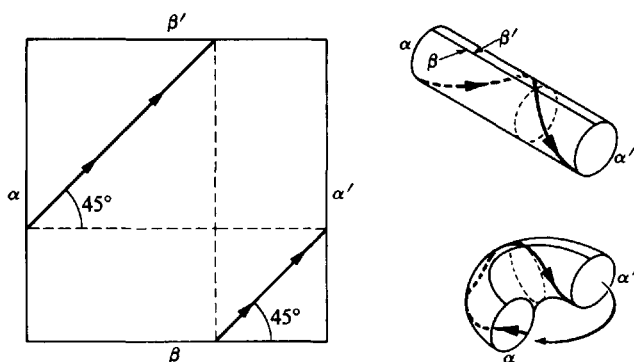


Figure 6.1

**Proposition 6.15.** Let  $A \in \mathcal{F}$  be a.s. invariant, then there is a set  $A' \in \mathcal{F}$  which is invariant such that  $P(A \triangle A') = 0$ .

*Proof.* Let

$$A'' = \bigcup_0^{\infty} T^{-n}A, \quad T^0A = A.$$

Then  $A'' = A$  a.s., and  $T^{-1}A'' \subset A''$ . Let

$$A' = \bigcap_0^{\infty} T^{-n}A'',$$

noting that  $A''$  is a.s. invariant gives  $A' = A$  a.s., and  $T^{-1}A'' \subset A''$  implies that  $A'$  is invariant.

The concept of ergodicity, as pointed out above, is a guarantee that the phase space does not split into parts of positive probability which are inaccessible to one another.

**Example.** Let  $\Omega$  be the unit square, that is,  $\Omega = \{(x, y); 0 \leq x < 1, 0 \leq y < 1\}$ ,  $\mathcal{F}$  the Borel field. Let  $a$  be any positive number and define  $T(x, y) = ((x + a)[1], (y + a)[1])$ . Use  $P = dx dy$ ; then it is easy to check that  $T$  is measure preserving. What we have done is to sew edges of  $\Omega$  together (see Fig. 6.1) so that  $\alpha$  and  $\alpha'$  are together,  $\beta$  and  $\beta'$ .  $T$  moves points at a  $45^\circ$  angle along the sewn-together square. Just by looking at Fig. 6.1 you can see that  $T$  does not move around the points of  $\Omega$  very much, and it is easy to construct invariant sets of any probability, for example, the shaded set as shown in Fig. 6.2.

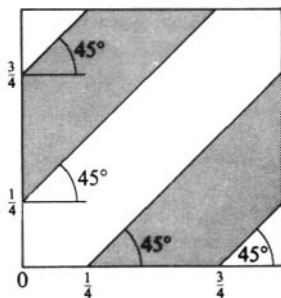


Figure 6.2

### Problems

5. In Problem 2(1), show that if  $\lambda$  is rational,  $T$  is not ergodic.

6. We use this problem to illustrate more fully the dynamical aspect. Take  $(\Omega, \mathcal{F})$  and let  $T: \Omega \rightarrow \Omega$  be measurable. Start with any point  $\omega \in \Omega$ , then  $\omega$  has the motion  $\omega \rightarrow T\omega \rightarrow T^2\omega \rightarrow \cdots$ . Let  $N_n(A, \omega)$  be the number of times that the moving point  $T^k\omega$  enters the set  $A \in \mathcal{F}$  during the first  $n$  motions; that is,  $N_n(A, \omega)$  is the number of times that  $T^k\omega \in A$ ,  $k = 0, 1, \dots, n-1$ . Keeping  $\omega$  fixed, define probabilities  $P_\omega^{(n)}(\cdot)$  on  $\mathcal{F}$  by  $P_\omega^{(n)}(\cdot) = N_n(\cdot, \omega)/n$ . That is,  $P_\omega^{(n)}(A)$  is the proportion of times in the first  $n$  moves that the point is in the set  $A$ . Let  $X$  be any random variable on  $(\Omega, \mathcal{F})$ .

a) Show that

$$\frac{1}{n} \sum_{k=0}^{n-1} X(T^k\omega) = \int X(\cdot) dP_\omega^{(n)}(\cdot).$$

Assume that there is a probability  $P_\omega(\cdot)$  on  $\mathcal{F}$  such that for every  $A \in \mathcal{F}$ ,  $\lim_n P_\omega^{(n)} = P_\omega(A)$ .

b) Show that  $T$  is  $P_\omega(\cdot)$ -preserving, that is,

$$P_\omega(T^{-1}A) = P_\omega(A), \quad \text{all } A \in \mathcal{F}.$$

c) Show that if  $X(\omega) \geq 0$  is bounded, then

$$\frac{1}{n} \sum_{k=0}^{n-1} X(T^k\omega) \rightarrow \int X(\cdot) dP_\omega(\cdot).$$

What is essential here is that the limit  $\int X dP_\omega$  not depend on where the system started at time zero. Otherwise, to determine the limiting time averages of (c) for the system, a detailed knowledge of the position  $\omega$  in phase-space at  $t = 0$  would be necessary. Hence, what we really need is the additional assumption that  $P_\omega(\cdot)$  be the same for all  $\omega$ , in other words, that the limiting proportion of time that is spent in the set  $A$  not depend on the starting position  $\omega$ . Now substitute this stronger assumption, that is: There is a probability  $P(\cdot)$  on  $\mathcal{F}$  such that for every  $A \in \mathcal{F}$  and  $\omega \in \Omega$ ,

$$\lim P_\omega^{(n)}(A) = P(A).$$

d) Show that under the above assumption,  $A$  invariant  $\Rightarrow A = \Omega$  or  $\emptyset$ .

This result shows not only that  $T$  is ergodic on  $(\Omega, \mathcal{F}, P)$ , but ergodic in a much stronger sense than that of Definition 6.14. The stronger assumption above is much too restrictive in the sense that most dynamical systems do not satisfy it. There are usually some starting states  $\omega$  which are exceptional in that the motion under  $T$  does not mix them up very well. Take, for example, elastic two-dimensional molecules in a rectangular box. At  $t = 0$  consider

the state  $\omega$  which is shown in Fig. 6.3, where all the molecules have the same  $x$ -coordinate and velocity. Obviously, there will be large chunks of phase-space that will never be entered if this is the starting state. What we want then, is some weaker version that says  $P_\omega^{(n)}(\cdot) \rightarrow P(\cdot)$  for most starting states  $\omega$ . With this weakening, the strong result of (d) above will no longer be true. We come back later to the appropriate weakening which, of course, will result in something like 6.14 instead of (d).

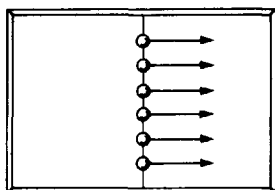


Figure 6.3

#### 4. INVARIANT RANDOM VARIABLES

Along with invariant sets go invariant random variables.

**Definition 6.16.** Let  $X(\omega)$  be a random variable on  $(\Omega, \mathcal{F}, P)$ ,  $T$  measure-preserving; then  $X(\omega)$  is called an invariant random variable if  $X(\omega) = X(T\omega)$ .

Note

**Proposition 6.17.**  $X$  is invariant iff  $X$  is measurable  $\mathcal{I}$ .

*Proof.* If  $X$  is invariant, then for every  $x$ ,  $\{X < x\} \in \mathcal{I}$ ; hence  $X$  is measurable  $\mathcal{I}$ . Conversely, if  $X(\omega) = \chi_A(\omega)$ ,  $A \in \mathcal{I}$ , then

$$X(T\omega) = \chi_A(T\omega) = \chi_{T^{-1}A}(\omega) = \chi_A(\omega).$$

Now consider the class  $\mathcal{L}$  of all random variables on  $(\Omega, \mathcal{I})$  which are invariant; clearly  $\mathcal{L}$  is closed under linear combinations, and  $X_n \in \mathcal{L}$ ,  $X_n(\omega) \uparrow X(\omega)$  implies

$$X(T\omega) = \lim_n X_n(T\omega) = \lim_n X_n(\omega) = X(\omega).$$

Hence by 2.38,  $\mathcal{L}$  contains all nonnegative random variables on  $(\Omega, \mathcal{I})$ . Thus clearly every random variable on  $(\Omega, \mathcal{I})$  is invariant.

The condition for ergodicity can be put very nicely in terms of invariant random variables.

**Proposition 6.18.** Let  $T$  be measure-preserving on  $(\Omega, \mathcal{F}, P)$ .  $T$  is ergodic iff every invariant random variable  $X(\omega)$  is a.s. equal to a constant.

*Proof.* One way is immediate; that is, for any invariant set  $A$ , let  $X(\omega) = \chi_A(\omega)$ . Then  $X(\omega)$  constant a.s. implies  $P(A) = 0, 1$ . Conversely, suppose  $P(X < x) = 0, 1$  for all  $x$ . Since for  $x \uparrow +\infty$ ,  $P(X < x) \rightarrow 1$ ,  $P(X < x) = 1$ , for all  $x$  sufficiently large. Let  $x_0 = \inf \{x; P(X < x) = 1\}$ . Then for

every  $\epsilon > 0$ ,  $P(x_0 - \epsilon < X < x_0 + \epsilon) = 1$ , and taking  $\epsilon \downarrow 0$  yields  $P(X = x_0) = 1$ .

Obviously we can weaken 6.18 to read

**Proposition 6.19.** *T is ergodic iff every bounded invariant random variable is a.s. constant.*

In general, it is usually difficult to show that a given transformation is ergodic. Various tricks are used: For example, we can apply 6.19 to Problem 2(1).

**Example.** We show that if  $\lambda$  is irrational, then  $Tx = (x + \lambda) [1]$  is ergodic. Let  $f(x)$  be any Borel-measurable function on  $[0, 1)$ . Assume it is in  $L_2(dx)$ , that is,  $\int f^2 dx < \infty$ . Then we have

$$f(x) = \sum_{-\infty}^{+\infty} c_n e^{2\pi i n x}, \quad \text{a.s.}(dx),$$

where the sum exists as a limit in the second mean, and  $\sum |c_n|^2 < \infty$ . Therefore

$$f(Tx) = \sum_{-\infty}^{+\infty} c_n e^{2\pi i n \lambda} \cdot e^{2\pi i n x}.$$

For  $f(x)$  to be invariant,  $c_n(1 - e^{2\pi i n \lambda}) = 0$ . This implies either  $c_n = 0$  or  $e^{2\pi i n \lambda} = 1$ . The latter can never be satisfied for nonzero  $n$  and irrational  $\lambda$ . The conclusion is that  $f(x) = c_0$  a.s.; by 6.19,  $T$  is ergodic.

## Problems

7. Use the method of the above example to show that the transformation of Problem 2(2) is ergodic.

8. Using 2.38, show that if  $T$  is measure-preserving on  $(\Omega, \mathcal{F}, P)$  and  $X(\omega)$  any random variable, that

$$(6.20) \quad E[X(\omega)] = E[X(T\omega)].$$

## 5. THE ERGODIC THEOREM

One of the most remarkable of the strong limit theorems is the result usually referred to as the ergodic theorem.

**Theorem 6.21.** *Let  $T$  be measure-preserving on  $(\Omega, \mathcal{F}, P)$ . Then for  $X$  any random variable such that  $E|X| < \infty$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} X(T^k \omega) = E(X | \mathcal{J}) \text{ a.s.}$$

To prove this result, we prove first an odd integration inequality:

**Theorem 6.22** (*Maximal ergodic theorem*). Let  $T$  be measure-preserving on  $(\Omega, \mathcal{F}, P)$ , and  $X$  a random variable such that  $E|X| < \infty$ . Define

$$S_k(\omega) = X(\omega) + \cdots + X(T^{k-1}\omega), \text{ and } M_n(\omega) = \max(0, S_1, S_2, \dots, S_n).$$

Then

$$\int_{\{M_n > 0\}} X dP \geq 0.$$

*Proof.* We give a very simple recent proof of this due to Adriano Garsia [61]. For any  $k \leq n$ ,  $M_n(T\omega) \geq S_k(T\omega)$ . Hence

$$X(\omega) + M_n(T\omega) \geq X(\omega) + S_k(T\omega) = S_{k+1}(\omega).$$

Write this as

$$X(\omega) \geq S_{k+1}(\omega) - M_n(T\omega), \quad k = 1, \dots, n.$$

But trivially,

$$X(\omega) \geq S_1(\omega) - M_n(T\omega),$$

since  $S_1(\omega) = X(\omega)$  and  $M_n(\omega) \geq 0$ . These two inequalities together give  $X(\omega) \geq \max(S_1(\omega), \dots, S_n(\omega)) - M_n(T\omega)$ . Thus

$$\int_{\{M_n > 0\}} X dP \geq \int_{\{M_n > 0\}} [\max(S_1(\omega), \dots, S_n(\omega)) - M_n(T\omega)] dP.$$

On the set  $\{M_n > 0\}$ ,  $\max(S_1, \dots, S_n) = M_n$ . Hence

$$\int_{\{M_n > 0\}} X dP \geq \int_{\{M_n > 0\}} [M_n(\omega) - M_n(T\omega)] dP,$$

but

$$\int_{\{M_n > 0\}} M_n(\omega) dP - \int_{\{M_n > 0\}} M_n(T\omega) dP \geq \int M_n(\omega) dP - \int M_n(T\omega) dP = 0.$$

This last is by (6.20).

*Completion of proof of 6.21.* Assuming that  $E(X | \mathcal{J}) = 0$ , prove that the averages converge to zero a.s. Then apply this result to the random variable  $X(\omega) - E(X | \mathcal{J})$  to get the general case. Let  $\bar{X} = \overline{\lim} S_n/n$ , and for any  $\epsilon > 0$ , denote  $D = \{\bar{X} > \epsilon\}$ . Note that  $\bar{X}(T\omega) = \bar{X}(\omega)$ , so  $\bar{X}$  and therefore  $D$  are invariant. Define the random variable

$$X^*(\omega) = (X(\omega) - \epsilon)\chi_D(\omega),$$

and using  $X^*$ , define  $S_k^*$ ,  $M_n^*$  as above.

The maximal ergodic theorem gives

$$\int_{\{M_n^* > 0\}} X^* dP \geq 0.$$

The rest of the proof is easy sailing. The sets

$$F_n = \{M_n^* > 0\} = \left\{ \max_{1 \leq k \leq n} S_k^* > 0 \right\}$$

converge upward to the set

$$\begin{aligned} F &= \left\{ \sup_{k \geq 1} S_k^* > 0 \right\} = \left\{ \sup_{k \geq 1} \frac{S_k^*}{k} > 0 \right\} \\ &= \left\{ \sup_{k \geq 1} \frac{S_k}{k} > \epsilon \right\} \cap D. \end{aligned}$$

Since  $\sup_{k \geq 1} S_k/k \geq \bar{X}$ ,  $F = D$ . The inequality  $E|X^*| \leq E|X| + \epsilon$  allows the use of the bounded convergence theorem, so we conclude that

$$\int_{F_n} X^* dP \rightarrow \int_F X^* dP.$$

Therefore,

$$\int_D X^* dP \geq 0.$$

But

$$\begin{aligned} \int_D X^* dP &= \int_D X dP - \epsilon P(D) \\ &= \int_D E(X | \mathfrak{J}) dP - \epsilon P(D) \\ &= -\epsilon P(D), \end{aligned}$$

which implies  $P(D) = 0$ , and  $\bar{X} \leq 0$  a.s. Apply the same argument to the random variable  $-X(\omega)$ . Here the lim sup of the sums is

$$\overline{\lim} \left( -\frac{S_n}{n} \right) = -\underline{\lim} \left( \frac{S_n}{n} \right) = -\underline{X}.$$

The conclusion above becomes  $-\underline{X} \leq 0$  or  $\underline{X} \geq 0$  a.s. Putting these two together gives the theorem. Q.E.D.

A consequence of 6.21 is that if  $T$  is ergodic, time averages can be replaced by phase averages, in other words,

**Corollary 6.23.** *Let  $T$  be measure-preserving and ergodic on  $(\Omega, \mathcal{F}, P)$ . Then for  $X$  any random variable such that  $E|X| < \infty$ ,*

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} X(T^k \omega) = EX \quad \text{a.s.}$$

*Proof.* Every set in  $\mathfrak{J}$  has probability zero or one, hence

$$E(X | \mathfrak{J}) = EX \quad \text{a.s.}$$



## 6. CONVERSES AND COROLLARIES

It is natural to ask whether the conditions of the ergodic theorem are necessary and sufficient. Again the answer is—Almost. If  $X$  is a non-negative random variable and  $EX = \infty$ , it is easy to show that for  $T$  measure-preserving and ergodic,

$$\frac{1}{n} \sum_0^{n-1} X(T^k \omega) \rightarrow \infty \quad \text{a.s.}$$

Because defining for  $\alpha > 0$ ,

$$X_\alpha(\omega) = \begin{cases} X(\omega), & \text{if } X(\omega) \leq \alpha, \\ 0, & \text{otherwise,} \end{cases}$$

of course,  $E|X_\alpha| < \infty$ . Thus the ergodic theorem can be used to get

$$\liminf \frac{1}{n} \sum_0^{n-1} X(T^k \omega) \geq \lim \frac{1}{n} \sum_0^{n-1} X_\alpha(T^k \omega) = EX_\alpha, \quad \text{a.s.}$$

Take  $\alpha \uparrow \infty$  to get the conclusion.

But, in general, if  $E|X| = \infty$ , it does not follow that the averages diverge a.s. (see Halmos [65, p. 32]).

Come back now to the question of the asymptotic density of the points  $\omega, T\omega, T^2\omega, \dots$ . In the ergodic theorem, for any  $A \in \mathcal{F}$ , take  $X(\omega) = \chi_A(\omega)$ . Then the conclusion reads, if  $T$  is ergodic,

$$\frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k \omega) \xrightarrow{\text{a.s.}} P(A),$$

so that for almost every starting point  $\omega$ , the asymptotic proportion of points in  $A$  is exactly  $P(A)$ . If  $\Omega$  has a topology with a countable basis such that  $P(N) > 0$  for every open neighborhood  $N$ , then this implies that for almost every  $\omega$ , the set of points  $\omega, T\omega, T^2\omega, \dots$  is dense in  $\Omega$ .

Another interesting and curious result is

**Corollary 6.24.** *Let  $T: \Omega \rightarrow \Omega$  be measure-preserving and ergodic with respect to both  $(\Omega, \mathcal{F}, P_1)$  and  $(\Omega, \mathcal{F}, P_2)$ . Then either  $P_1 = P_2$  or  $P_1$  and  $P_2$  are orthogonal in the sense that there is a set  $A \in \mathcal{J}$  such that  $P_1(A) = 1, P_2(A^c) = 1$ .*

*Proof.* If  $P_1 \neq P_2$ , take  $B \in \mathcal{F}$  such that  $P_1(B) \neq P_2(B)$  and let  $X(\omega) = \chi_B(\omega)$ . Let  $A$  be the set of  $\omega$  such that

$$\frac{1}{n} \sum_0^{n-1} X(T^k \omega) \rightarrow P_1(B).$$

By the ergodic theorem  $P_1(A) = 1$ . But  $A^c$  includes all  $\omega$  such that

$$\frac{1}{n} \sum_0^{n-1} X(T^k \omega) \rightarrow P_2(B),$$

and we see that  $P_2(A^c) = 1$ .

Finally, we ask concerning convergence in the first mean of the averages to  $EX$ . By the Lebesgue theorem we know that a.s. convergence plus just a little more gives first mean convergence. But here we have to work a bit to get the additional piece.

**Corollary 6.25.** *Under the conditions of the ergodic theorem*

$$\lim_n E \left| \frac{1}{n} \sum_{k=0}^{n-1} X(T^k \omega) - E(X | \mathcal{I}) \right| = 0.$$

*Proof.* We can assume that  $E(X | \mathcal{I}) = 0$ . Let

$$V_n = \frac{1}{n} \sum_{k=0}^{n-1} X(T^k \omega).$$

Since  $V_n \xrightarrow{\text{a.s.}} 0$ , by Egoroff's theorem, for any  $\epsilon > 0$ ,  $\exists A \in \mathcal{F}$  such that  $P(A) \leq \epsilon$  and  $V_n \rightarrow 0$  uniformly on  $A^c$ . Now,

$$\overline{\lim}_n E |V_n| = \overline{\lim}_n \int_A |V_n| dP \leq \overline{\lim}_n \frac{1}{n} \sum_{k=0}^{n-1} \int_A |X_k| dP, \quad X_k = X(T^k \omega).$$

These integrals can be estimated by

$$\begin{aligned} \int_A |X_k| dP &= \int_{A \cap \{|X_k| > N\}} |X_k| dP + \int_{A \cap \{|X_k| \leq N\}} |X_k| dP \\ &\leq \int_{A \cap \{|X_k| > N\}} |X_k| dP + NP(A) \leq \int_{\{|X| > N\}} |X| dP + NP(A). \end{aligned}$$

Since  $\epsilon$  is arbitrary, conclude that for any  $N$ ,

$$\overline{\lim}_n E |V_n| \leq \int_{\{|X| > N\}} |X| dP.$$

Let  $N$  go to infinity; then by the bounded convergence theorem, the right-hand side above goes to zero.

## Problems

**9.** Another consequence of the ergodic theorem is a weak form of Weyl's equidistribution theorem. For any  $x$  in  $[0, 1)$  and interval  $I \subset [0, 1)$ , let  $R_x^{(n)}(I)$  be the proportion of the points  $\{(x + \lambda)[1], (x + 2\lambda)[1], \dots, (x + n\lambda)[1]\}$  falling in the interval  $I$ . If  $\lambda \geq 0$  is irrational, show that for  $x$  in a set of Lebesgue measure one,  $R_x^{(n)}(I) \rightarrow \text{length } I$ .

**10.** Let  $\mu$  be a finite measure on  $\mathcal{B}_1([0, 1))$  such that  $Tx = 2x[1]$  preserves  $\mu$ -measure. Show that  $\mu$  is singular with respect to Lebesgue measure.

11. Let  $T$  be measurable on  $(\Omega, \mathcal{F})$ , and define  $\mathcal{M}$  as the set of all probabilities  $P$  on  $\mathcal{F}$  such that  $T$  is measure-preserving on  $(\Omega, \mathcal{F}, P)$ . Define real linear combinations by  $(\alpha P_1 + \beta P_2)(B) = \alpha P_1(B) + \beta P_2(B)$ ,  $B \in \mathcal{F}$ . Show that

a)  $\mathcal{M}$  is convex, that is, for  $\alpha, \beta \geq 0$ ,  $\alpha + \beta = 1$ ,

$$P_1, P_2 \in \mathcal{M} \Rightarrow \alpha P_1 + \beta P_2 \in \mathcal{M}.$$

An extreme point of  $\mathcal{M}$  is a probability  $P \in \mathcal{M}$  which is not a linear combination  $\alpha P_1 + \beta P_2$ ,  $\alpha, \beta > 0$ ,  $\alpha + \beta = 1$ , with  $P_1, P_2 \in \mathcal{M}$ . Show that

b) the extreme points of  $\mathcal{M}$  are the probabilities  $P$  such that  $T$  is ergodic on  $(\Omega, \mathcal{F}, P)$ .

## 7. BACK TO STATIONARY PROCESSES

By the ergodic theorem and its corollary, if the shift-transformation  $S$  (see 6.10) is ergodic on  $(R^{(\infty)}, \mathcal{B}_\infty, \hat{P})$ , then

$$\frac{1}{n} \sum_{k=1}^n \hat{X}_k \rightarrow EX_1 \quad \text{a.s. and first mean.}$$

If  $S$  is ergodic, then, the same conclusions will hold for the original  $X_1, X_2, \dots$  process, because a.s. convergence and  $r$ th mean convergence depend only on the distribution of the process.

Almost all the material concerning invariance and ergodicity, can be formulated in terms of the original process  $X_1, X_2, \dots$  rather than going into representation space. If  $B \in \mathcal{B}_\infty$  and  $A = \{X \in B\}$ , then the inverse image under  $X$  of  $S^{-1}B$ ,  $S$  the shift operator, is

$$\{X \in S^{-1}B\} = \{\omega; (X_2, X_3, \dots) \in B\}.$$

Hence, we reach

**Definition 6.26.** An event  $A \in \mathcal{F}$  is invariant if  $\exists B \in \mathcal{B}_\infty$  such that for every  $n \geq 1$

$$A = \{(X_n, X_{n+1}, \dots) \in B\}.$$

The class of invariant events is easily seen to be a  $\sigma$ -field. Similarly, we define a random variable  $Z$  to be invariant if there is a random variable  $\varphi$  on  $(R^{(\infty)}, \mathcal{B}_\infty)$  such that

$$(6.27) \quad Z = \varphi(X_n, X_{n+1}, \dots), \quad \text{all } n \geq 1.$$

The results of Section 4 hold again;  $Z$  is invariant iff it is  $\mathcal{I}$ -measurable. The ergodic theorem translates as

**Theorem 6.28.** If  $X_1, X_2, \dots$  is a stationary process,  $\mathcal{I}$  the  $\sigma$ -field of invariant events, and  $E|X_1| < \infty$ , then

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{a.s.}} E(X_1 | \mathcal{I}).$$

*Proof.* From the ergodic theorem,  $S_n/n$  converges a.s. to some random variable  $Y$ . It is not difficult to give an argument that the correct translation is  $Y = E(X_1 | \mathfrak{I})$ . But we can identify  $Y$  directly: take  $Y = \overline{\lim} S_n/n$ , then the sets  $\{Y < y\}$  are invariant, hence  $Y$  is  $\mathfrak{I}$ -measurable. Take  $A \in \mathfrak{I}$ , then since we have first mean convergence,

$$(6.29) \quad \frac{1}{n} \sum_{k=1}^n \int_A X_k dP \rightarrow \int_A Y dP.$$

Select  $B \in \mathcal{B}_\infty$  so that  $A = \{(X_k, \dots) \in B\}$ , for all  $k \geq 1$ . Now stationarity gives

$$\int_{\{(X_k, \dots) \in B\}} X_k dP = \int_{\{(X_1, \dots) \in B\}} X_1 dP = \int_A X_1 dP.$$

Use this in 6.29, to conclude

$$\int_A X_1 dP = \int_A Y dP, \quad \text{all } A \in \mathfrak{I}.$$

By definition,  $Y = E(X_1 | \mathfrak{I})$ .

**Definition 6.30.** A stationary process  $X_1, X_2, \dots$  is ergodic if every invariant event has probability zero or one.

If  $\mathfrak{I}$  has this zero-one property, of course the averages converge to  $EX_1$ , a.s.

Ergodicity, like stationarity, is preserved under taking functions of the process. More precisely,

**Proposition 6.31.** Let  $X_1, X_2, \dots$  be a stationary and ergodic process,  $\varphi(\mathbf{x})$  measurable  $\mathcal{B}_\infty$ , then the process  $Y_1, Y_2, \dots$  defined by

$$Y_k = \varphi(X_k, X_{k+1}, \dots)$$

is ergodic.

*Proof.* This is very easy to see. Use the same argument as in Proposition 6.6 to conclude that for any  $B \in \mathcal{B}_\infty$ ,  $\exists A \in \mathcal{B}_\infty$  such that

$$\{\omega; (Y_1, Y_2, \dots) \in B\} = \{\omega; (X_1, X_2, \dots) \in A\},$$

$$\{\omega; (Y_2, Y_3, \dots) \in B\} = \{\omega; (X_2, X_3, \dots) \in A\}.$$

Hence, every invariant event on the  $Y$ -process coincides with an invariant event on the  $X$ -process.

One result that is both interesting and useful in establishing ergodicity is

**Proposition 6.32.** Let  $X_1, X_2, \dots$  be a stationary process. Then every invariant event  $A$  is a tail event.

*Proof.* Take  $B$  so that  $A = \{(X_n, X_{n+1}, \dots) \in B\}$ ,  $n \geq 1$ . Hence  $A \in \mathcal{F}(X_n, X_{n+1}, \dots)$ , all  $n$ .

**Corollary 6.33.** *Let  $X_1, X_2, \dots$  be independent and identically distributed; then the process is ergodic.*

*Proof.* Kolmogorov's zero-one law.

By this corollary we can include the strong law of large numbers as a consequence of the ergodic theorem, except for the converse.

### Problems

12. Show that the event  $\{X_n \in B \text{ i.o.}\}$ ,  $B \in \mathcal{B}_1$ , is invariant.

13. If  $X_1, X_2, \dots$  is stationary and if there  $\exists B \in \mathcal{B}_\infty$  such that

$$A = \{(X_1, X_2, \dots) \in B\} = \{(X_2, X_3, \dots) \in B\} \quad \text{a.s.,}$$

show that  $A$  is a.s. equal to an invariant event.

14. A process  $X_1, X_2, \dots$  is called a normal process if  $X_1, \dots, X_n$  have a joint normal distribution for every  $n$ . Let  $EX_i = 0$ ,  $\Gamma_{ij} = EX_i X_j$ .

1) Prove that the process is stationary iff  $\Gamma_{ij}$  depends only on  $|i - j|$ .

2) Assume  $\Gamma_{ij} = r(|i - j|)$ . Then prove that  $\lim_m r(m) = 0$  implies that the process is ergodic.

[Assume that for every  $n$ , the determinant of  $\Gamma_{ij}$ ,  $i, j = 1, \dots, n$ , is not zero. See Chapter 11 for the definition and properties of joint normal distributions.]

15. Show that  $X_1, X_2, \dots$  is ergodic iff for every  $A \in \mathcal{B}_k$ ,  $k = 1, 2, \dots$ ,

$$\frac{1}{n} \sum_{k=1}^n \chi_A(X_n, \dots, X_{n+k}) \xrightarrow{\text{a.s.}} P((X_1, \dots, X_{k+1}) \in A).$$

16. Let  $X_1, X_2, \dots$  and  $Y_1, Y_2, \dots$  be two stationary, ergodic processes on  $(\Omega, \mathcal{F}, P)$ . Toss a coin with probability  $p$  of heads independently of  $\mathbf{X}$  and  $\mathbf{Y}$ . If it comes up heads, observe  $\mathbf{X}$ , if tails, observe  $\mathbf{Y}$ . Show that the resultant process is stationary, but not ergodic.

### 8. AN APPLICATION

There is a very elegant application of the above ideas due to Spitzer, Kesten, and Whitman (see Spitzer, [130, pp. 35 ff]). Let  $X_1, X_2, \dots$  be a sequence of independent identically distributed random variables taking values in the integers. The range  $R_n$  of the first  $n$  sums is defined as the number of distinct points in the set  $\{S_1, \dots, S_n\}$ . Heuristically, the more the tendency of the sums  $S_n$  to return to the origin, the smaller  $R_n$  will be, because if we are at a given point  $k$  at time  $n$ , the distribution of points around  $k$  henceforth looks like the distribution of points around the origin starting from  $n = 1$ . To

pin this down, write

$$P(\text{no return}) = P(S_n \neq 0, \quad n = 1, 2, \dots),$$

then,

**Proposition 6.34**

$$\lim_n \frac{ER_n}{n} = P(\text{no return}).$$

*Proof.* Write  $R_n = \sum_{k=0}^n W_k$ , where

$$W_k = \begin{cases} 1 & \text{if } S_j \neq S_k, \quad j = 1, \dots, k-1, \\ 0, & \text{otherwise.} \end{cases}$$

So now,

$$\begin{aligned} EW_k &= P(S_k - S_{k-1} \neq 0, S_k - S_{k-2} \neq 0, \dots, S_k - S_1 \neq 0) \\ &= P(X_k \neq 0, X_k + X_{k+1} \neq 0, \dots, X_k + \dots + X_2 \neq 0) \\ &= P(S_1 \neq 0, \dots, S_{k-1} \neq 0), \end{aligned}$$

the last equality holding because  $(X_k, \dots, X_2)$  has the same distribution as  $(X_1, \dots, X_{k-1})$ . Therefore  $\lim_k EW_k = P(\text{no return})$ .

The remarkable result is

**Theorem 6.35**

$$\lim_n \frac{R_n}{n} = P(\text{no return}) \quad \text{a.s.}$$

*Proof.* Take  $N$  any positive integer, and let  $Z_k$  be the number of distinct points visited by the successive sums during the time  $(k-1)N+1$  to  $kN$ , that is,  $Z_k$  is the range of  $\{S_{(k-1)N+1}, \dots, S_{kN}\}$ . Note that  $Z_k$  depends only on the  $X_n$  for  $n$  between  $(k-1)N+1$  and  $kN$ , so that the  $Z_k$  are independent,  $|Z_k| \leq N$ , and are easily seen to be identically distributed. Use the obvious inequality  $R_{nN} \leq Z_1 + \dots + Z_n$  and apply the law of large numbers:

$$\overline{\lim}_n \frac{R_{nN}}{nN} \leq \overline{\lim}_n \frac{1}{nN} (Z_1 + \dots + Z_n) = \frac{1}{N} EZ_1.$$

For  $n'$  not a multiple of  $N$ ,  $R_{n'}$  differs by at most  $N$  from one of  $R_{nN}$ , so

$$\overline{\lim}_n \frac{R_n}{n} \leq \frac{1}{N} EZ_1 \quad \text{a.s.}$$

But  $Z_1 = R_N$ , hence letting  $N \rightarrow \infty$ , and using 6.34, we get

$$\overline{\lim}_n \frac{R_n}{n} \leq P(\text{no return}) \quad \text{a.s.}$$

Going the other way is more interesting. Define

$$V_k = \begin{cases} 1 & \text{if } S_j \neq S_k, \quad j > k, \\ 0, & \text{otherwise.} \end{cases}$$

That is,  $V_k$  is one if at time  $k$ ,  $S_k$  is in a state which is never visited again, and zero otherwise. Now  $V_1 + \cdots + V_n$  is the number of states visited in time  $n$  which are never revisited.  $R_n$  is the number of states visited in time  $n$  which are not revisited prior to time  $n + 1$ . Thus  $R_n \geq V_1 + \cdots + V_n$ . Now define

$$\varphi(X_1, X_2, \dots) = \begin{cases} 1, & S_k \neq 0, \quad k = 1, \dots, \\ 0, & \text{otherwise,} \end{cases}$$

and make the important observation that

$$\begin{aligned} V_k &= \begin{cases} 1, & X_k \neq 0, X_k + X_{k+1} \neq 0, \dots, \\ 0, & \text{otherwise,} \end{cases} \\ &= \varphi(X_k, X_{k+1}, \dots). \end{aligned}$$

Use 6.31 and 6.33 to show that  $V_1, V_2, \dots$  is a stationary, ergodic sequence. Now the ergodic theorem can be used to conclude that

$$\liminf \frac{R_n}{n} \geq \liminf \frac{V_1 + \cdots + V_n}{n} = EV_1 \quad \text{a.s.}$$

Of course,  $EV_1 = P(\text{no return})$ , and this completes the proof.

## 9. RECURRENCE TIMES

For  $X_0, X_1, \dots$  stationary, and any set  $A \in \mathcal{B}_1$ , look at the times that the process enters  $A$ , that is, the  $n$  such that  $X_n \in A$ .

**Definition 6.36.** For  $A \in \mathcal{B}_1$ ,  $P(X_0 \in A) > 0$ , define

$$\begin{aligned} R_1 &= \min \{n; X_n \in A, n > 0\}, \\ R_2 &= \min \{n; X_n \in A, n > R_1\}, \end{aligned}$$

and so forth. These are the occurrence times of the set  $A$ . The recurrence times  $T_1, T_2, \dots$  are given by

$$(6.37) \quad T_1 = R_1, \quad T_2 = R_2 - R_1, \quad T_k = R_k - R_{k-1}.$$

If  $\{X_n\}$  is ergodic, then  $P(X_n \in A \text{ i.o.}) = 1$ ; so the  $R_k$  are well defined a.s. But if  $\{X_n\}$  is not ergodic a subsidiary condition has to be imposed to make the  $R_k$  well defined. At any rate, the smaller  $A$  is, the longer it takes to get

back to it, and in fact we arrive at the following proposition:

**Proposition 6.38.** *Let  $X_0, X_1, \dots$  be a stationary process,  $A \in \mathcal{B}_1$  such that*

$$(6.39) \quad P(X_n \in A \text{ at least once}) = 1,$$

*then the  $R_k, k = 1, \dots$  are finite a.s. On the sample space  $\Omega_A = \{\omega; X_0 \in A\}$  the  $T_1, T_2, \dots$  form a stationary sequence under the probability  $P(\cdot | X_0 \in A)$ , and*

$$E(T_1 | X_0 \in A) = \frac{1}{P(X_0 \in A)}.$$

**Remarks.** This means that to get the  $T_1, T_2, \dots$  to be stationary, we have to start off on the set  $A$  at time zero. This seems too complicated because once we have landed in  $A$  then the returns should be stationary, that is, the  $T_2, T_3, \dots$  should be stationary under  $P$ . This is not so, and counterexamples are not difficult to construct (see Problem 20.)

Note that  $P(X_0 \in A) > 0$ , otherwise condition (6.39) is violated. Therefore, conditional probability given  $\{X_0 \in A\}$  is well defined.

*Proof.* Extend  $\{X_n\}$  to  $0, \pm 1, \pm 2, \dots$ . By (6.39),  $P(R_1 < \infty) = 1$ . From the stationarity

$$P(T_k \leq n, R_{k-1} = j) = P(R_1 \leq n, C),$$

where  $C \in \mathcal{F}(X_0, X_{-1}, \dots)$ . Let  $n \rightarrow \infty$ , so that we get

$$P(T_k < \infty, R_{k-1} = j) = P(R_{k-1} = j).$$

Go down the ladder to conclude that  $P(R_1 < \infty) = 1$  implies  $P(R_k < \infty) = 1, k \geq 1$ .

To prove stationarity, we need to establish that

$$\begin{aligned} P(T_1 = n_1, \dots, T_k = n_k | X_0 \in A) \\ = P(T_2 = n_1, \dots, T_{k+1} = n_k | X_0 \in A). \end{aligned}$$

This is not difficult to do, but to keep out of notational messes, I prove only that

$$P(T_1 = n | X_0 \in A) = P(T_2 = n | X_0 \in A).$$

The generalization is exactly the same argument. Define random variables  $U_n = \chi_A(X_n)$ , and sets  $C_k$  by

$$C_k = \{U_{-1} = 0, \dots, U_{-k+1} = 0, U_{-k} = 1\}.$$

The  $\{C_k\}$  are disjoint, and

$$\bigcup_1^\infty C_k = \{X_n \in A \text{ at least once, } n \leq -1\}.$$



I assert that  $P\left(\bigcup_1^{\infty} C_k\right) = 1$ , because

$$P\left(\bigcup_1^{\infty} C_k\right) = \lim_n P\left(\bigcup_{-n-1}^{-1} \{U_k = 1\}\right).$$

By stationarity of the  $\{U_k\}$  process,

$$P\left(\bigcup_{-n-1}^{-1} \{U_k = 1\}\right) = P\left(\bigcup_0^n \{U_k = 1\}\right).$$

The limit of the right-hand side is  $P(X_n \in A \text{ at least once})$ , which is one by (6.39). Now, using stationarity again, we find that

$$\begin{aligned} P(T_2 = n, X_0 \in A) &= \sum_{k=1}^{\infty} P(T_2 = n, T_1 = k, X_0 \in A) \\ &= \sum_{k=1}^{\infty} P(T_1 = n, X_0 \in A, C_k) = P(T_1 = n, X_0 \in A). \end{aligned}$$

To compute  $E(T_1 | X_0 \in A)$ , note that

$$P(T_1 \geq k, X_0 \in A) = P(C_k).$$

Consequently,

$$\begin{aligned} E(T_1 | X_0 \in A) &= \sum_{k=1}^{\infty} P(T_1 \geq k | X_0 \in A) \\ &= \frac{1}{P(X_0 \in A)} \sum_1^{\infty} P(C_k) = \frac{1}{P(X_0 \in A)}. \end{aligned}$$

We can use the ergodic theorem to get a stronger result.

**Theorem 6.40.** *If the process  $\{X_n\}$  is ergodic, then the process  $\{T_k\}$  on  $\{X_0 \in A\}$  is ergodic under  $P(\cdot | X_0 \in A)$ .*

*Proof.* By the ergodic theorem,

$$\frac{1}{n} \sum_{k=1}^n \chi_A(X_k) \xrightarrow{\text{a.s.}} P(X_0 \in A).$$

On every sequence such that the latter holds,  $R_n \rightarrow \infty$ , and

$$\frac{1}{R_n} \sum_{k=1}^{R_n} \chi_A(X_k) \rightarrow P(X_0 \in A).$$

The sum  $\sum_{k=1}^{R_n} \chi_A(X_k)$  is the number of visits  $X_k$  makes to  $A$  up to the time of the  $n$ th occurrence. Thus  $\sum_1^{R_n} \chi_A(X_k) = n$ , so

$$\frac{T_1 + \cdots + T_n}{n} \xrightarrow{\text{a.s.}} \frac{1}{P(X_0 \in A)}.$$

Note that for every function  $f$  measurable  $\mathcal{B}_\infty$ , there is a function  $g$  measurable  $\mathcal{B}_\infty(\{0, 1\})$  such that on the set  $\{R_{k-1} = j\}$ ,

$$f(T_k, T_{k+1}, \dots) = g(U_{j+1}, U_{j+2}, \dots),$$

where  $U_n$  denotes  $\chi_A(X_n)$  again. Therefore

$$\sum_{k=1}^n f(T_k, \dots) = \sum_{j=0}^{R_{n-1}} U_j g(U_{j+1}, \dots).$$

Since  $R_{n-1} \rightarrow \infty$  a.s., if  $E|g| < \infty$  we can use the ergodic theorem as follows:

$$\begin{aligned} (6.41) \quad \frac{1}{n} \sum_{k=1}^n f(T_k, \dots) &= \frac{R_{n-1}}{n} \cdot \frac{1}{R_{n-1}} \sum_{j=0}^{R_{n-1}} U_j g(U_{j+1}, \dots) \\ &\rightarrow \frac{1}{P(X_0 \in A)} E(U_0 g(U_1, \dots)) \quad \text{a.s.} \end{aligned}$$

Because  $U_0 = \chi_A(X_0)$ , this limit is

$$E(g(U_1, \dots) | X_0 \in A) \quad \text{or} \quad E(f(T_1, \dots) | X_0 \in A).$$

On the space  $\{X_0 \in A\}$ , take  $f$  to be any bounded invariant function of  $T_1, \dots$ , that is,

$$f(T_1, T_2, \dots) = f(T_k, T_{k+1}, \dots), \quad \text{all } k \geq 1.$$

Then (6.41) implies that  $f(T_1, \dots)$  is a.s. constant on  $\{X_0 \in A\}$ , implying in turn that the process  $T_1, T_2, \dots$  is ergodic.

## 10. STATIONARY POINT PROCESSES

Consider a class of processes gotten as follows: to every integer  $n$ , positive and negative, associate a random variable  $U_n$  which is either zero or one. A way to look at this process is as a sequence of points. If a point occurs at time  $n$ , then  $U_n = 1$ , otherwise  $U_n = 0$ . We impose a condition to ensure that these processes have a.s. no trivial sample points.

**Condition 6.42.** *There is probability zero that all  $U_n = 0$ .*

For a point process to be time-homogeneous, the  $\{U_n\}$  have to form a stationary sequence.

**Definition 6.43.** *A stationary discrete point process is a stationary process  $\{U_n\}$ ,  $n = 0, \pm 1, \dots$  where  $U_n$  is either zero or one.*

*Note:*  $P(U_n = 1) > 0$ , otherwise  $P(\text{all } U_n = 0) = 1$ . Take  $A = \{1\}$ , then (6.38) implies that given  $\{U_0 = 1\}$  the times between points form a stationary

sequence  $T_1, T_2, \dots$  of positive integer-valued random variables such that

$$ET_1 = \frac{1}{P(U_0 = 1)} \quad \text{or} \quad P(U_0 = 1) = \frac{1}{ET_1}.$$

The converse question is: Given a stationary sequence  $T_1, T_2, \dots$  of positive integer valued random variables, is there a stationary point process  $\{U_n\}$  with interpoint distances having the same distribution as  $T_1, T_2, \dots$ ? The difficulty in extracting the interpoint distances from the point process was that an origin had to be pinned down; that is,  $\{U_0 = 1\}$  had to be given. But here the problem is to start with the  $T_1, T_2, \dots$  and smear out the origin somehow. Define

$$V_k = \begin{cases} 1, & \exists n \text{ such that } T_1 + \dots + T_n = k, \\ 0, & \text{otherwise.} \end{cases}$$

Suppose that there is a stationary process  $\{U_k\}$  with interpoint distances having the same distribution as  $T_1, \dots$ . Denote the probability on  $\{U_n\}$  by  $Q$ . Then for  $s$  any sequence  $k$ -long of zeroes and ones

$$Q((U_k, \dots, U_1) = s \mid U_0 = 1) = P((V_k, \dots, V_1) = s).$$

This leads to

(6.44)

$$\begin{aligned} Q((U_k, \dots, U_1) = s) &= \sum_{j=0}^{\infty} Q((U_k, \dots, U_1) = s, U_0 = 0, \dots, 0, U_{-j} = 1) \\ &= \sum_{j=0}^{\infty} Q((U_{k+j}, \dots, U_{j+1}) = s, U_j = 0, \dots, U_0 = 1) \\ &= Q(U_0 = 1) \sum_{j=0}^{\infty} P((V_{k+j}, \dots, V_{j+1}) = s, T_1 > j) \\ &= \frac{1}{ET_1} \sum_{j=0}^{\infty} P((V_{k+j}, \dots, V_{j+1}) = s, T_1 > j). \end{aligned}$$

The right-hand side above depends only upon the probability  $P$  on the  $T_1, T_2, \dots$  process. So if a stationary point process exists with interpoint distances  $T_1, T_2, \dots$  it must be unique. Furthermore, if we define  $Q$  directly by means of (6.44), then it is not difficult to show that we get a consistent set of probabilities for a stationary point process having interpoint distances with distribution  $T_1, T_2, \dots$ . Thus

**Theorem 6.45.** *Let  $T_1, T_2, \dots$  be a stationary sequence of positive integer valued random variables such that  $ET_1 < \infty$ . Then there is a unique stationary point process  $\{U_n\}$  such that the interpoint distances given  $\{U_0 = 1\}$  have the same distribution as  $T_1, T_2, \dots$*

*Proof.* The work is in showing that the  $Q$  defined in (6.44) is a probability for the desired process. To get stationarity one needs to verify that

$$Q((U_{k+1}, \dots, U_2) = s) = Q((U_k, \dots, U_1) = s).$$

Once this is done, then it is necessary to check that the interpoint distances given  $\{U_0 = 1\}$  have the same distribution as  $T_1, T_2, \dots$ . For example, check that

$$Q(U_j = 1, U_{j-1} = 0, \dots, U_1 = 0 \mid U_0 = 1) = P(T_1 = j).$$

All this verification we leave to the reader.

### Problems

17. Given the process  $T_1, T_2, \dots$  such that  $T_k = 2$ , all  $k$ , describe the corresponding stationary point process.

18. If the interpoint distances for a stationary point process, given  $\{U_0 = 1\}$  are  $T_1, T_2, \dots$ , prove that the distribution of time  $n^*$  until the first point past the origin, that is,

$$n^* = \min \{n; n \geq 1, U_n = 1\}$$

is given by

$$P(n^* = n) = \frac{1}{ET_1} P(T_1 > n).$$

19. If the interpoint distances  $T_1, T_2, \dots$  given  $\{U_0 = 1\}$  are independent random variables, then show that the unconditioned interpoint distances  $T_2, T_3, \dots$  are independent identically distributed random variables with the same distribution as the conditioned random variables and that  $(T_2, T_3, \dots)$  is independent of the time  $n^*$  of the first point past the origin.

20. Consider the stationary point process  $\{U_n\}$  having independent interpoint distances  $T_1, T_2, \dots$  with  $P(T_1 = 1) = 1 - \epsilon$ , with  $\epsilon$  very small. Now consider the stationary point process

$$V_n = 1 - U_n, \quad n = 0, \pm 1, \dots$$

Show that for this latter process the random variables defined by the distance from the first point past the origin to the second point, from the second point to the third point, etc., do not form a stationary sequence.

### NOTES

The ergodic theorem was proven by G. D. Birkhoff in 1931 [4]. Since that time significant improvements have been made on the original lengthy proof, and the proof given here is the most recent simplification in a sequence involving many simplifications and refinements.

The nicest texts around on ergodic theorems in a framework of measure-preserving transformations are Halmos [65], and Hopf [73] (the latter available only in German). From the point of view of dynamical systems, see Khintchine [90], and Birkhoff [5]. Recently, E. Hopf generalized the ergodic theorem to operators  $T$  acting on measurable functions defined on a measure space  $(\Omega, \mathcal{F}, \mu)$  such that  $T$  does not increase  $L_1$ -norm or  $L_\infty$ -norm. For this and other operator-theoretic aspects, see the Dunford-Schwartz book [41].

For  $X_1, \dots$  a process on  $(\Omega, \mathcal{F}, P)$ , we could try to define a set transformation  $S^{-1}$  on  $\mathcal{F}(X)$  similar to the shift in sequence space as follows: If  $A \in \mathcal{F}(X)$ , then take  $B \in \mathcal{B}_\infty$  such that  $A = \{(X_1, X_2, \dots) \in B\}$  and define  $S^{-1}A = \{(X_2, X_3, \dots) \in B\}$ . The same difficulty comes up again;  $B$  is not uniquely determined, and if  $A = \{X \in B_1\}$ ,  $A = \{X \in B_2\}$ , it is not true in general that

$$\{(X_2, X_3, \dots) \in B_1\} = \{(X_2, X_3, \dots) \in B_2\}.$$

But for stationary processes, it is easy to see that these latter two sets differ only by a set of probability zero. Therefore  $S^{-1}$  can be defined, not on  $\mathcal{F}(X)$ , but only on equivalence classes of sets in  $\mathcal{F}(X)$ , where sets  $A_1, A_2 \in \mathcal{F}(X)$  are equivalent if  $P(A_1 \Delta A_2) = 0$ . For a deeper discussion of this and other topics relating to the translation into a probabilistic context of the ergodic theorem see Doob [39].

The fact that the expected recurrence time of  $A$  starting from  $A$  is  $1/P(X_0 \in A)$  is due to Kac [81]. A good development of point processes and proofs of a more general version of 6.45 and the ergodic property of the recurrence times is in Ryll-Nardzewski [119].

## CHAPTER 7

# MARKOV CHAINS

### 1. DEFINITIONS

The basic property characterizing Markov chains is a probabilistic analogue of a familiar property of dynamical systems. If one has a system of particles and the position and velocities of all particles are given at time  $t$ , the equations of motion can be completely solved for the future development of the system. Therefore, any other information given concerning the past of the process up to time  $t$  is superfluous as far as future development is concerned. The present state of the system contains all relevant information concerning the future. Probabilistically, we formalize this by defining a Markov chain as

**Definition 7.1.** A process  $X_0, X_1, \dots$  taking values in  $F \in \mathcal{B}_1$  is called a Markov chain if

$$P(X_{n+1} \in A \mid X_n, \dots, X_0) = P(X_{n+1} \in A \mid X_n) \text{ a.s.}$$

for all  $n \geq 0$  and  $A \in \mathcal{B}_1(F)$ .

For each  $n$ , there is a version  $p_n(A \mid x)$  of  $P(X_{n+1} \in A \mid X_n = x)$  which is a regular conditional distribution. These are the *transition probabilities* of the process. We restrict ourselves to the study of a class of Markov chains which have a property similar to conservative dynamical systems. One way to state that a system is conservative is that if it goes from any state  $x$  at  $t = 0$  to  $y$  in time  $\tau$ , then starting from  $x$  at any time  $t$  it will be in  $y$  at time  $t + \tau$ . The corresponding property for Markov chains is that the transition probabilities do not depend on time.

**Definition 7.2.** A Markov chain  $X_0, X_1, \dots$  on  $F \in \mathcal{B}_1$  has stationary transition probabilities  $p(A \mid x)$  if  $p(A \mid x)$  is a regular conditional distribution and if for each  $A \in \mathcal{B}_1(F)$ ,  $n \geq 0$ ,  $p(A \mid x)$  is a version of  $P(X_{n+1} \in A \mid X_n = x)$ . The initial distribution is defined by

$$\pi(A) = P(X_0 \in A).$$

The transition probabilities and the initial distribution determine the distribution of the process.

**Proposition 7.3.** For a Markov chain  $X_0, X_1, \dots$  on  $F \in \mathcal{B}_1$ ,

(7.4)

$$P(X_n \in A_n, X_{n-1} \in A_{n-1}, \dots, X_0 \in A_0) \\ = \int_{A_{n-1}} \int_{A_{n-2}} \cdots \int_{A_0} p_{n-1}(A_n | x_{n-1}) p_{n-2}(dx_{n-1} | x_{n-2}) \cdots p_0(dx_1 | x_0) \pi(dx_0)$$

for all  $A_n, \dots, A_0 \in \mathcal{B}_1(F)$ .

*Proof.* Use

$$P(X_n \in A_n, X_{n-1} \in A_{n-1}, \dots, X_0 \in A_0) \\ = \int_{\{X_{n-1} \in A_{n-1}, \dots, X_0 \in A_0\}} P(X_n \in A_n | X_{n-1}, \dots, X_0) dP$$

and proceed by induction.

The Markov property as defined in 7.1 simply states that the present state of the system determines the probability for one step into the future. This generalizes easily:

**Proposition 7.5.** Let  $X_0, X_1, \dots$  be a Markov chain, and  $C$  any event in  $\mathcal{F}(X_{n+1}, X_{n+2}, \dots)$ . Then

$$P(C | X_n, \dots, X_0) = P(C | X_n) \quad \text{a.s.}$$

Having stationary transition probabilities generalizes into

**Proposition 7.6.** If the process  $X_0, X_1, \dots$  is Markov on  $F \in \mathcal{B}_1$  with stationary transition probabilities, then for every  $B \in \mathcal{B}_\infty(F)$  there are versions of

$$P((X_{n+1}, X_{n+2}, \dots) \in B | X_n = x)$$

and

$$P((X_1, X_2, \dots) \in B | X_0 = x),$$

which are equal.

The proofs of both 7.5 and 7.6 are left to the reader.

Proposition 7.3 indicates how to do the following construction:

**Proposition 7.7.** Let  $p(A | x)$  be a regular conditional probability on  $\mathcal{B}_1(F)$  for  $x \in F$ , and  $\pi(A)$  a probability on  $\mathcal{B}_1(F)$ . Then there exists a Markov chain  $\hat{X}_0, \hat{X}_1, \hat{X}_2, \dots$  with stationary transition probabilities  $p(A | x)$  and initial distribution  $\pi$ .

*Proof.* Use (7.4) to define probabilities on rectangles. Then it is easy to check that all the conditions of 2.18 are met. Now extend and use the coordinate representation process. What remains is to show that any process satisfying (7.4) is a Markov chain with  $p(A | x)$  as its transition

probabilities. (That  $P(X_0 \in A) = \pi(A)$  is obvious.) For this, use (7.4) for  $n - 1$  to conclude that (7.4) can be written as

$$\begin{aligned} P(X_n \in A_n, \dots) &= \int_{A_{n-1}} p(A_n | x_{n-1}) P(X_{n-1} \in dx_{n-1}, X_{n-2} \in A_{n-2}, \dots) \\ &= \int_{\{x_{n-1} \in A_{n-1}, x_{n-2} \in A_{n-2}, \dots\}} p(A_n | x_{n-1}) P(X_{n-1} \in dx_{n-1}, \\ &\quad X_{n-2} \in dx_{n-2}, \dots). \end{aligned}$$

But, by definition,

$$P(X_n \in A_n, \dots) = \int_{\{X_{n-1} \in A_{n-1}, \dots\}} P(X_n \in A_n | X_{n-1}, \dots) dP.$$

Hence  $p(A | x)$  is a version of  $P(X_n \in A | X_{n-1} = x_{n-1}, \dots)$  for all  $n$ .

Now we switch points of view. For the rest of this chapter forget about the original probability space  $(\Omega, \mathcal{F}, P)$ . Fix one version  $p(A | x)$  of the stationary transition probabilities and consider  $p(A | x)$  to be the given data. Each initial distribution  $\pi$ , together with  $p(A | x)$ , determines a probability on  $(F^{(\infty)}, \mathcal{B}_\infty(F))$  we denote by  $P_\pi$ , and the corresponding expectation by  $E_\pi$ . Under  $P_\pi$ , the coordinate representation process  $\hat{X}_0, \hat{X}_1, \dots$  becomes a Markov chain with initial distribution  $\pi$ . Thus we are concerned with a family of processes having the same transition probability, but different initial distributions. If  $\pi$  is concentrated on the single point  $\{x\}$ , then denote the corresponding probability on  $\mathcal{B}_\infty(F)$  by  $P_x$ , and the expectation by  $E_x$ . Under  $P_x$ ,  $\hat{X}_0, \hat{X}_1, \dots$  is referred to as the "process starting from the point  $x$ ." Now eliminate the  $\wedge$  over variables, with the understanding that henceforth all processes referred to are coordinate representation processes with  $P$  one of the family of probabilities  $\{P_\pi\}$ . For any  $\pi$ , and  $B \in \mathcal{B}_\infty(F)$ , always use the version of  $P_\pi((X_0, \dots) \in B | X_0 = x)$  given by  $P_x((X_0, \dots) \in B)$ .

In exactly the same way as in Chapter 3, call a nonnegative integer-valued random variable  $n^*$  a *stopping time* or *Markov time* for a Markov chain  $X_0, X_1, \dots$  if

$$\{n^* = n\} \in \mathcal{F}(X_0, \dots, X_n).$$

Define  $\mathcal{F}(X_n, n \leq n^*)$  to be the  $\sigma$ -field consisting of all events  $A$  such that  $A \cap \{n^* \leq n\} \in \mathcal{F}(X_0, \dots, X_n)$ . Then

**Proposition 7.8.** *If  $n^*$  is a Markov time for  $X_0, X_1, \dots$ , then*

$$P((X_{n^*}, X_{n^*+1}, \dots) \in B | X_n, n \leq n^*) = P_{X_{n^*}}((X_0, X_1, \dots) \in B) \quad \text{a.s. } P.$$

*Note.* The correct interpretation of the above is: Let

$$\varphi(x) = P_x((X_0, X_1, \dots) \in B).$$

Then the right-hand side is  $\varphi(X_{n^*})$ .

Proposition 7.8 is called the *strong Markov property*.



*Proof.* Let  $A \in \mathcal{F}(X_n, n \leq n^*)$ , then the integral of the left-hand side of 7.8 over  $A$  is

$$\begin{aligned} P((X_n^*, X_{n^*+1}^*, \dots) \in B, A) &= \sum_{n=0}^{\infty} P((X_n^*, X_{n^*+1}^*, \dots) \in B, A, n^* = n) \\ &= \sum_{n=0}^{\infty} P((X_n, X_{n+1}, \dots) \in B, A, n^* = n). \end{aligned}$$

The set  $A \cap \{n^* = n\} \in \mathcal{F}(X_0, \dots, X_n)$ . Hence

$$\begin{aligned} P((X_n, \dots) \in B, A, n^* = n) &= \int_{A \cap \{n^* = n\}} P((X_n, \dots) \in B \mid X_n, \dots) dP \\ &= \int_{A \cap \{n^* = n\}} \varphi(X_n) dP. \end{aligned}$$

Putting this back in, we find that

$$\begin{aligned} P_x((X_n^*, \dots) \in B, A) &= \sum_{n=0}^{\infty} \int_{A \cap \{n^* = n\}} \varphi(X_n) dP \\ &= \int_A \varphi(X_{n^*}) dP. \end{aligned}$$

A special case of a Markov chain are the successive sums  $S_0, S_1, S_2, \dots$  of independent random variables  $Y_1, Y_2, \dots$  ( $S_0 = 0$  convention). This is true because independence gives

$$P(S_{n+1} \in A \mid S_n, \dots, S_0) = P(S_{n+1} \in A \mid S_n) \quad \text{a.s.}$$

If the summands are identically distributed, then the chain has stationary transition probabilities:

$$\begin{aligned} P(S_{n+1} \in A \mid S_n = x) &= P(Y_{n+1} \in A - x) \\ &= F(A - x) \quad \text{a.s.,} \end{aligned}$$

where  $F(B)$  denotes the probability  $P(Y_1 \in B)$  on  $\mathcal{B}_1$ . In this case, take  $F(A - x)$  as the fixed conditional probability distribution to be used. Now letting  $X_0$  have any initial distribution  $\pi$  and using the transition probability  $F(A - x)$  we get a Markov chain  $X_0, X_1, \dots$  having the same distribution as  $Y_0, Y_0 + Y_1, Y_0 + Y_1 + Y_2, \dots$  where  $Y_0, Y_1, \dots$  are independent,  $Y_0$  has the distribution  $\pi$ , and  $Y_1, Y_2, \dots$  all have the distribution  $F$ . In particular, the process "starting from  $x$ " has the distribution of  $S_0 + x, S_1 + x, S_2 + x, \dots$ . Call any such Markov process a *random walk*.

## Problems

1. Define the  $n$ -step transition probabilities  $p^{(n)}(A \mid x)$  for all  $A \in \mathcal{B}_1(F)$ ,  $x \in F$  by

$$p^{(1)}(A \mid x) = p(A \mid x), \quad p^{(n+1)}(A \mid x) = \int p^{(n)}(A \mid y) p(dy \mid x).$$

a) Show that  $p^{(n)}(A | x)$  equals  $P_x(X_n \in A)$ , hence is a version of

$$P(X_n \in A | X_0 = x).$$

b) Show that  $p^{(n)}(A | x)$  is a regular conditional probability on  $\mathcal{B}_1(F)$  given  $x \in F$ , and that for all  $A \in \mathcal{B}_1(F)$ ,  $x \in F$ ,  $n, m > 0$ ,

$$p^{(n+m)}(A | x) = \int p^{(n)}(A | y) p^{(m)}(dy | x).$$

2. Let  $\varphi(x)$  be a bounded  $\mathcal{B}_\infty$ -measurable function. Demonstrate that  $E_x \varphi(X_1, X_2, \dots)$  is  $\mathcal{B}_1$ -measurable in  $x$ .

## 2. ASYMPTOTIC STATIONARITY

There is a class of limit theorems which state that certain processes are asymptotically stationary. Generally these theorems are formulated as: Given a process  $X_1, X_2, \dots$ , a stationary process  $X_1^*, X_2^*, \dots$  exists such that for every  $B \in \mathcal{B}_\infty$ ,

$$\lim_n P((X_n, X_{n+1}, \dots) \in B) = P^*((X_1^*, X_2^*, \dots) \in B).$$

The most well-known of these relate to Markov chains with stationary transition probabilities  $p(A | x)$ . Actually, what we would really like to show for Markov chains is that no matter what the initial distribution of  $X_0$  is, convergence toward the same stationary limiting distribution sets in, that is, that for all  $B \in \mathcal{B}_\infty(F)$ , and initial distributions  $\pi$ ,

$$\lim_n P_\pi((X_n, \dots) \in B) = P^*((X_1^*, \dots) \in B),$$

where  $\{X_n^*\}$  is stationary.

Suppose, to begin, that there is a limiting distribution  $\bar{\pi}(\cdot)$  on  $\mathcal{B}_1(F)$  such that for all  $x \in F$ ,  $A \in \mathcal{B}_1(F)$ ,

$$\lim_n P_x(X_n \in A) = \bar{\pi}(A).$$

If this limiting distribution  $\bar{\pi}(A)$  exists, then from

$$\begin{aligned} P_\pi(X_n \in A) &= \int P(X_n \in A | X_0 = x) \pi(dx) \\ &= \int P_x(X_n \in A) \pi(dx) \end{aligned}$$

comes

$$P_\pi(X_n \in A) \rightarrow \bar{\pi}(A), \quad \text{all } \pi.$$

Also,

$$P_\pi(X_{n+1} \in A) = \int P(X_{n+1} \in A | X_n = x) P_\pi(X_n \in dx).$$

For  $A$  fixed, approximate  $p(A | x)$  uniformly by simple functions of  $x$ . Taking limits implies that  $\bar{\pi}(A)$  must satisfy

$$(7.9) \quad \bar{\pi}(A) = \int p(A | x) \bar{\pi}(dx).$$

If  $\bar{\pi}$  is the limiting distribution of  $X_n$ , what happens if we start the process off with the distribution  $\bar{\pi}$ ? The idea is that if  $\bar{\pi}$  is a limiting steady-state distribution, then, starting the system with this distribution, it should maintain a stable behavior. This is certainly true in the following sense—let us start the process with any initial distribution satisfying (7.9). Then this distribution maintains itself throughout time, that is,

$$P(X_n \in A) = P(X_{n-1} \in A) = \cdots = P(X_0 \in A).$$

This is established by iterating (7.9) to get  $\bar{\pi}(A) = \int p^{(n)}(A | x) \bar{\pi}(dx)$ . In this sense any solution of (7.9) gives stable initial conditions to the process.

**Definition 7.10.** For transition probabilities  $p(A | x)$ , an initial distribution  $\pi(A)$  will be called a stationary initial distribution if it satisfies (7.9).

But if a stationary initial distribution is used for the process, much more is true.

**Proposition 7.11.** Let  $X_0, X_1, \dots$  be a Markov process with stationary transition probabilities such that the initial distribution  $\pi(A)$  satisfies (7.9); then the process is stationary.

*Proof.* By 7.6 there are versions of  $P(X_n \in A_n, \dots, X_1 \in A_1 | X_0 = x)$ ,  $P(X_{n+1} \in A_n, \dots, X_2 \in A_1 | X_1 = x)$  which are equal. Since  $P(X_1 \in A) = P(X_0 \in A) = \pi(A)$ , integrating these versions over  $A_0$  gives

$$P(X_n \in A_n, \dots, X_0 \in A_0) = P(X_{n+1} \in A_n, \dots, X_1 \in A_0),$$

which is sufficient to prove the process stationary.

Furthermore, Markov chains have the additional property that if  $P_x(X_n \in A)$  converges to  $\bar{\pi}(A)$ , all  $A \in \mathcal{B}_1(F)$ ,  $x \in F$ , then this one-dimensional convergence implies that the distribution of the entire process is converging to the distribution of the stationary process with initial distribution  $\bar{\pi}$ .

**Proposition 7.12.** If  $p^{(n)}(A | x) \rightarrow \bar{\pi}(A)$ , all  $A \in \mathcal{B}_1(F)$ ,  $x \in F$ , then for any  $B \in \mathcal{B}_\infty(F)$ , and all  $x \in F$ ,

$$P_x((X_{n+1}, \dots) \in B) \rightarrow P_{\bar{\pi}}((X_1, X_2, \dots) \in B).$$

*Proof.* Write

$$P_x((X_1, X_2, \dots) \in B) = \varphi(x).$$

Then

$$P_x((X_{n+1}, \dots) \in B | X_n = y) = \varphi(y), \quad \text{a.s.}$$

and, also,

$$P_{\bar{\pi}}((X_1, X_2, \dots) \in B \mid X_0 = x) = \varphi(x), \quad \text{a.s.}$$

Now,

$$\begin{aligned} P((X_{n+1}, \dots) \in B \mid X_0 = x) \\ = \int P((X_{n+1}, \dots) \in B \mid X_n = y) P(X_n \in dy \mid X_0 = x) \end{aligned}$$

or

$$P((X_{n+1}, \dots) \in B \mid X_0 = x) = \int \varphi(y) p^{(n)}(dy \mid x).$$

Under the stated conditions, one can show, by taking simple functions that approximate  $\varphi(x)$  uniformly, that

$$\int \varphi(y) p^{(n)}(dy \mid x) \rightarrow \int \varphi(y) \bar{\pi}(dy) = P_{\bar{\pi}}((X_1, \dots) \in B).$$

Therefore, the asymptotic behavior problem becomes: How many stationary initial distributions does a given set of transition probabilities have, and does  $p^{(n)}(A \mid x)$  converge to some stationary distribution as  $n \rightarrow \infty$ ?

### 3. CLOSED SETS, INDECOMPOSABILITY, ERGODICITY

**Definition 7.13.** A set  $A \in \mathcal{B}_1(F)$  is closed if  $p(A \mid x) = 1$  for all  $x \in A$ .

The reason for this definition is obvious; if  $X_0 \in A$ , and  $A$  is closed, then  $X_n \in A$  with probability one for all  $n$ . Hence if there are two closed disjoint sets  $A_1, A_2$ , then

$$p^{(n)}(A_1 \mid x) = \begin{cases} 1, & x \in A_1, \quad n \geq 0, \\ 0, & x \in A_2, \quad n \geq 0, \end{cases}$$

and there is no hope that  $p^{(n)}(A_1 \mid x)$  converges to the same limit for all starting points  $x$ .

**Definition 7.14.** A chain is called indecomposable if there are no two disjoint closed sets  $A_1, A_2 \in \mathcal{B}_1(F)$ .

Use a stationary initial distribution  $\bar{\pi}$ , if one exists, to get the stationary process  $X_0, X_1, \dots$ . If the process is in addition ergodic, then use the ergodic theorem to assert

$$(7.15) \quad \frac{1}{n} \sum_{k=1}^n \chi_A(X_k) \rightarrow \bar{\pi}(A) \quad \text{a.s.}$$

Take conditional expectations of (7.15), given  $X_0 = x$ . Use the boundedness and proposition 4.24, to get

$$\frac{1}{n} \sum_{k=1}^n p^{(k)}(A \mid x) \rightarrow \bar{\pi}(A), \quad \text{a.s. } \bar{\pi}(dx).$$

Thus, from ergodicity, it is possible to get convergence of the averages of the  $p^{(n)}(A | x)$  to  $\bar{\pi}(dx)$  a.s.  $\bar{\pi}(dx)$ . The questions of ergodicity of the process uniqueness of stationary initial distributions, and indecomposability go together.

**Theorem 7.16.** *Let the chain be indecomposable. Then if a stationary initial distribution  $\bar{\pi}$  exists, it is unique and the process gotten by using  $\bar{\pi}$  as the initial distribution is ergodic.*

*Proof.* Let  $C$  be an invariant event under the shift transformation in  $\mathcal{B}_\infty$ ,  $\bar{\pi}$  a stationary initial distribution. Take  $\varphi(x) = P_x(C)$ . Now, using  $\bar{\pi}$  as the initial distribution, write

$$P(C | X_0) = E(P(C | X_1, X_0) | X_0).$$

By the Markov property, since  $C \in \mathcal{F}(X_n, X_{n+1}, \dots)$  for all  $n \geq 0$ ,

$$P(C | X_1, X_0) = P(C | X_1), \quad \text{a.s.}$$

By Proposition 7.6,

$$P(C | X_0 = x) = P(S^{-1}C | X_1 = x) \quad \text{a.s. } \bar{\pi},$$

and by the invariance of  $C$ , then

$$P(C | X_1 = x) = \varphi(x).$$

Therefore  $\varphi(x)$  satisfies

$$(7.17) \quad \varphi(x) = \int \varphi(y)p(dy | x) \quad \text{a.s. } \bar{\pi}(dx).$$

By a similar argument,  $P(C | X_n, \dots, X_0) = P(C | X_n) = \varphi(X_n)$  a.s. and  $E(\varphi(X_n) | X_{n-1}, \dots, X_0) = \varphi(X_{n-1})$  a.s. Apply the martingale theorem to get  $\varphi(X_n) \rightarrow \chi_C$  a.s. Thus for any  $\epsilon > 0$ ,

$$\bar{\pi}(x; \epsilon \leq \varphi(x) \leq 1 - \epsilon) = 0$$

because the distribution of  $X_n$  is  $\bar{\pi}$ . So  $\varphi(x)$  can assume only the two values 0, 1 a.s.  $\bar{\pi}(dx)$ . Define sets

$$A_1 = \{x; \varphi(x) = 1\} \quad A_2 = \{x; \varphi(x) = 0\}.$$

Since  $\varphi(x)$  is a solution of (7.17),

$$1 = \int \varphi(y)p(dy | x), \quad x \in A_1,$$

$$0 = \int \varphi(y)p(dy | x), \quad x \in A_2,$$

except for a set  $D$ , such that  $\bar{\pi}(D) = 0$ . Therefore

$$p(A_i | x) = 1, \quad x \in A_i - D, \quad i = 1, 2.$$

Let us define

$$A_i^{(0)} = A_i$$

$$A_i^{(1)} = A_i - D$$

$$A_i^{(n+1)} = A_i^{(n)} \cap \{x; p(A_i^{(n)} | x) = 1\}.$$

If  $p(A_i^{(n)} | x) = 1$ ,  $x \in A_i^{(n)}$ , a.s.  $\bar{\pi}$ , then  $\bar{\pi}(A_i^{(n)} - A_i^{(n+1)}) = 0$ . Take  $C_i = \bigcap_0^\infty A_i^{(n)}$ . Then  $\bar{\pi}(C_i) = \bar{\pi}(A_i)$ , but the  $C_i$  are closed and disjoint. Hence one of  $C_1, C_2$  is empty,  $\varphi(x)$  is zero a.s. or one a.s.,  $P(C) = 0$  or 1, and the process is ergodic. Now, suppose there are two stationary initial distributions  $\bar{\pi}_1$  and  $\bar{\pi}_2$  leading to probabilities  $P_1$  and  $P_2$  on  $(\Omega, \mathcal{F})$ . By 6.24 there is an invariant  $C$  such that  $P_1(C) = 1$ , but  $P_2(C) = 0$ . Using the stationary initial distribution  $\bar{\pi} = \frac{1}{2}\bar{\pi}_1 + \frac{1}{2}\bar{\pi}_2$  we get the probability

$$P = \frac{1}{2}P_1 + \frac{1}{2}P_2$$

which is again ergodic, by the above argument. But

$$P(C) = \frac{1}{2}P_1(C) + \frac{1}{2}P_2(C) = \frac{1}{2}. \quad \text{Q.E.D.}$$

What has been left is the problem: When does there exist a  $\bar{\pi}(A)$  such that  $p^{(n)}(A | x) \rightarrow \bar{\pi}(A)$ ? If  $F$  is countable, this problem has a complete solution. In the case of general state spaces  $F$ , it is difficult to arrive at satisfactory conditions (see Doob, Chapter 6). But if a stationary initial distribution  $\bar{\pi}(A)$  exists, then under the following conditions:

- 1) the state space  $F$  is indecomposable under  $p(A | x)$ ;
- 2) the motion is *nonperiodic*; that is,  $F$  is indecomposable under the transition probabilities  $p^{(n)}(A | x)$ ,  $n = 2, 3, \dots$ ,
- 3) for each  $x \in F$ ,  $p(A | x) \ll \bar{\pi}(A)$ ;

Doob [35] has shown that

**Theorem 7.18.**  $\lim_n p^{(n)}(A | x) = \bar{\pi}(A)$  for all  $A \in \mathcal{B}_1(F)$ ,  $x \in F$ .

The proof is essentially an application of the ergodic theorem and its refinements. [As shown in Doob's paper, (3) can be weakened somewhat.]

#### 4. THE COUNTABLE CASE

The case where the state space  $F$  is countable is much easier to understand and analyze. It also gives some insight into the behavior of general Markov chains. Hence assume that  $F$  is a subset of the integers, that we have transition probabilities  $p(j | k)$ , satisfying

$$(7.19) \quad p(j | k) \geq 0, \quad \sum_j p(j | k) = 1,$$

where the summation is over all states in  $F$ , and  $n$ -step transition probabilities  $p^{(n)}(j|k)$  defined by

$$p^{(n+1)}(j|k) = \sum_i p^{(n)}(j|i)p(i|k).$$

This is exactly matrix multiplication: Denote the matrix  $\{p(j|k)\}$  by  $\mathbf{P}$ ; then the  $n$ -step transition probabilities are the elements of  $\mathbf{P}^n$ . Therefore, if  $F$  has a finite number of states the asymptotic stationarity problem can be studied in terms of what happens to the elements of matrix as it is raised to higher and higher powers. The theory in this case is complete and detailed. (See Feller [59, Vol. I, Chapter 16].) The idea that simplifies the theory in the countable case is the renewal concept. That is, if a Markov chain starts in state  $j$ , then every time it returns to state  $j$ , the whole process starts over again as from the beginning.

## 5. THE RENEWAL PROCESS OF A STATE

Let  $X_0, X_1, \dots$  be the Markov chain starting from state  $j$ . We ignore transitions to all other states and focus attention only on the returns of the process to the state  $j$ . Define random variables  $U_1, U_2, \dots$  by

$$U_n = \begin{cases} 1 & \text{if } X_n = j, \\ 0, & \text{otherwise.} \end{cases}$$

By the Markov property and stationarity of the transition probabilities, for any  $B \in \mathcal{B}_\infty(\{0, 1\})$ ,  $(s_1, \dots, s_{n-1}) \in \{0, 1\}^{(n-1)}$ ,

$$(7.20) \quad P_j((U_{n+1}, \dots) \in B \mid U_n = 1, U_{n-1} = s_{n-1}, \dots) = P_j((U_1, \dots) \in B).$$

This simple relationship partially summarizes the fact that once the process returns to  $j$ , the process starts anew. In general, any process taking values in  $\{0, 1\}$  and satisfying (7.20) is called a *renewal process*. We study the behavior at a single state  $j$  by looking at the associated process  $U_1, U_2, \dots$  governing returns to  $j$ . Define the event  $G$  that a return to state  $j$  occurs at least once by  $G = \bigcup_{n \geq 1} \{X_n = j\}$ .

**Theorem 7.21.** *The following dichotomy is in force*

$$P_j(G) = 1 \Rightarrow \begin{cases} P_j(X_n = j \text{ i.o.}) = 1, \\ \sum_1^\infty P_j(X_n = j) = \infty. \end{cases}$$

$$P_j(G) < 1 \Rightarrow \begin{cases} P_j(X_n = j \text{ i.o.}) = 0, \\ \sum_1^\infty P_j(X_n = j) < \infty. \end{cases}$$

*Proof.* Let  $F_n = \{U_n = 1, U_{n+k} = 0, \text{ all } k \geq 1\}$ ,  $n \geq 1$ ;  $F_0 = \{U_k = 0, \text{ all } k \geq 1\}$ . Thus  $F_n$  is the event that the last return to  $j$  occurs at time  $n$ . Hence

$$\{X_n = j \text{ i.o.}\}^c = \bigcup_{n \geq 0} F_n.$$

The  $F_n$  are disjoint, so

$$1 - P_j(X_n = j \text{ i.o.}) = \sum_0^\infty P_j(F_n),$$

and by (7.20),

$$\begin{aligned} (7.22) \quad P_j(F_n) &= P_j(U_{n+k} = 0, \text{ all } k \geq 1 \mid U_n = 1)P_j(U_n = 1) \\ &= P_j(F_0)P_j(U_n = 1). \end{aligned}$$

According to the definitions,  $F_0 = G^c$ , hence

$$P_j(G) = 1 \Rightarrow P_j(F_0) = 0 \Rightarrow P_j(F_n) = 0,$$

for all  $n \geq 0$ ; so  $P_j(X_n = j \text{ i.o.}) = 1$ . Then

$$\sum_1^\infty P_j(X_n = j) = \infty;$$

otherwise, the Borel-Cantelli lemma would imply  $P_j(X_n = j \text{ i.o.}) = 0$ . If  $P_j(G) < 1$ , then  $P_j(F_0) > 0$ , and we can use expression (7.22) to substitute for  $P_j(F_n)$ , getting

$$(7.23) \quad 1 - P_j(X_n = j \text{ i.o.}) = P_j(F_0) \left( 1 + \sum_1^\infty P_j(X_n = j) \right).$$

This implies  $\sum_1^\infty P_j(X_n = j) < \infty$  and thus  $P_j(X_n = j \text{ i.o.}) = 0$ .

**Definition 7.24.** Call the state  $j$  recurrent if  $P_j(X_n = j \text{ i.o.}) = 1$ , transient if  $P_j(X_n = j \text{ i.o.}) = 0$ .

Note that 7.21 in terms of transition probabilities reads, “ $j$  is recurrent iff  $\sum_1^\infty p^{(n)}(j \mid j) = \infty$ .”

Define  $R_k$  as the time of the  $n$ th return to  $j$ , and the times between returns as  $T_1 = R_1$ ,  $T_k = R_k - R_{k-1}$ ,  $k > 1$ . Then

**Proposition 7.25.** If  $j$  is recurrent, then  $T_1, T_2, \dots$  are independent, identically distributed random variables under the probability  $P_j$ .

*Proof.*  $T_1$  is a Markov time for  $X_0, X_1, \dots$ . By the strong Markov property 7.8,

$$P_j((X_{T_1+1}, X_{T_1+2}, \dots) \in B \mid X_n, n \leq T_1) = P_j((X_1, X_2, \dots) \in B).$$

Therefore, the process  $(X_{T_1+1}, \dots)$  has the same distribution as  $(X_1, \dots)$  and is independent of  $\mathcal{F}(X_n, n \leq T_1)$ , hence is independent of  $T_1$ .



The  $T_1, T_2, \dots$  are also called the *recurrence times* for the state  $j$ . The result of 7.25 obviously holds for any renewal process with  $T_1, T_2, \dots$ , the times between successive ones in the  $U_1, U_2, \dots$  sequence.

**Definition 7.26.** Call the recurrent state  $j$

positive-recurrent if  $E_j T_1 < \infty$ , null-recurrent if  $E_j T_1 = \infty$ .

**Definition 7.27.** The state  $j$  has period  $d \geq 1$  if  $T_1$  is distributed on the lattice  $L_d$ ,  $d \geq 1$  under  $P_j$ . If  $d > 1$ , call the state periodic; if  $d = 1$ , nonperiodic. [Recall from (3.32) that  $L_d = \{nd\}$ ,  $n = 0, \pm 1, \dots$ ]

For  $j$  recurrent, let the random vectors  $Z_k$  be defined by

$$Z_k = (X_{R_k}, \dots, X_{R_{k+1}-1}).$$

Then  $Z_k$  takes values in the space  $R$  of all finite sequences of integers with  $\mathcal{B}$  the smallest  $\sigma$ -field containing all sets of the form

$$\{(x_1, \dots, x_m) \in R; x_k = i\}, k \leq m,$$

where  $i$  is any integer. Since the length of blocks is now variable, an interesting generalization of 7.25 is

**Theorem 7.28.** The  $Z_0, Z_1, \dots$  are independent and identically distributed random vectors under the probability  $P_j$ .

*Proof.* This follows from 7.8 by seeing that  $T_1$  is a Markov time.  $Z_0$  is measurable  $\mathcal{F}(X_n, n \leq T_1)$  because for  $B \in \mathcal{B}$ ,

$$\{Z_0 \in B\} \cap \{T_1 = k\} = \{(X_0, \dots, X_{k-1}) \in B, T_1 = k\} \in \mathcal{F}(X_0, \dots, X_n),$$

$$k \leq n.$$

By 7.8, for  $A \in \mathcal{B}_\infty$ ,

$$P_j((X_{T_1}, X_{T_1+1}, \dots) \in A, Z_0 \in B) = P_j((X_0, X_1, \dots) \in A)P_j(Z_0 \in B).$$

Now  $Z_1$  is the same function of the  $X_{T_1}, X_{T_1+1}, \dots$  process as  $Z_0$  is of the  $X_0, X_1, \dots$  process. Hence  $Z_1$  is independent of  $Z_0$  and has the same distribution.

Call events  $\{A_n\}$  a *renewal event* & if the random variables  $\chi_{A_n}$  form a renewal process. Problems 3 and 4 are concerned with these.

## Problems

3. (Runs of length at least  $N$ ). Consider coin-tossing (biased or fair) and define

$$A_n = \{\omega_n = T, \omega_{n-1} = H, \dots, \omega_{n-N} = H\}.$$

Prove that  $\{A_n\}$  form a renewal event. [Note that  $A_n$  is the event such that at time  $n$  a run of at least  $N$  heads has just finished.]

4. Given any sequence  $t_N$ ,  $N$  long of  $H, T$ , let  $(\Omega, \mathcal{F}, P)$  be the coin-tossing game (fair or biased). Define  $A_n = \{\omega; (\omega_n, \dots, \omega_{n-N+1}) = t_N\}$ ,

( $A_n = \emptyset$  if  $n < N$ ). Find necessary and sufficient conditions on  $t_N$  for the  $\{A_n\}$  to form a renewal event.

5. If  $P_j(X_n = j \text{ i.o.}) = 0$ , then show that

$$P_j(X_n = j \text{ never occurs}) = \frac{1}{1 + \sum_{i=1}^{\infty} P_j(X_n = j)}.$$

6. Use Problem 5 to show that for a biased coin

$$P(\text{no equalizations}) = |p - q|.$$

7. Use Theorem 7.21 to show that if  $\{Z_n\}$  are the successive fortunes in a fair coin-tossing game,

$$P(Z_n = 0 \text{ i.o.}) = 1.$$

## 6. GROUP PROPERTIES OF STATES

**Definition 7.29.** If there is an  $n$  such that  $p^{(n)}(k | j) > 0$ ,  $j$  to  $k$  (denoted by  $j \rightarrow k$ ) is a permissible transition. If  $j \rightarrow k$  and  $k \rightarrow j$ , say that  $j$  and  $k$  communicate, and write  $j \leftrightarrow k$ .

Communicating states share properties:

**Theorem 7.30.** If  $j \leftrightarrow k$ , then  $j$  and  $k$  are simultaneously transient, null-recurrent, or positive-recurrent, and have the same period.

*Proof.* Use the martingale result, Problem 9, Chapter 5, to deduce that under any initial distribution the sets  $\{X_n = j \text{ i.o.}\}$  and  $\{X_n = k \text{ i.o.}\}$  have the same probability. So both are recurrent or transient together. Let  $T_1, T_2, \dots$  be the recurrence times for  $j$  starting from  $j$ , and assume  $E_j T_1 < \infty$ . Let  $V_n = 1$  or  $0$  as there was or was not a visit to state  $k$  between times  $R_{n-1}$  and  $R_n$ . The  $V_n$  are independent and identically distributed (7.28) with  $P(V_n = 1) > 0$ . Denote by  $n^*$  the first  $n$  such that  $V_n = 1$ ,  $T$  the time of the first visit to state  $k$ . Then

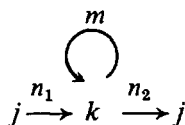
$$E_j T \leq \sum_{n=1}^{\infty} \int_{\{n^* = n\}} R_n dP_j = \sum_{n=1}^{\infty} \int_{\{n^* \geq n\}} T_n dP_j.$$

But  $\{n^* \geq n\} = \{n^* < n\}^c$ , and  $\{n^* < n\} \in \mathcal{F}(X_k, k \leq R_{n-1})$ , thus is independent of  $T_n$ . Hence

$$E_j T \leq E_j T_1 \cdot \sum_{n=1}^{\infty} P_j(n^* \geq n) = E_j T_1 \cdot E_j n^*.$$

Obviously,  $E_j n^* < \infty$ , so  $E_j T < \infty$ . Once  $k$  has occurred, the time until another occurrence of  $k$  is less than the time to get back to state  $j$  plus the time until another occurrence of  $k$  starting from  $j$ . This latter time has the same distribution as  $T$ . The former time must have finite expectation, otherwise  $E_j T_1 = \infty$ . Hence  $k$  has a recurrence time with finite expectation.

Take  $n_1, n_2$  such that  $p^{(n_1)}(k | j) > 0$ ,  $p^{(n_2)}(j | k) > 0$ . If  $T^{(k)}$  is the first return time to  $k$ , define  $I_k = \{n; P_k(T^{(k)} = n) > 0\}$ ,  $L_{d_1}$  is the smallest lattice containing  $I_k$ ,  $L_{d_2}$  the smallest lattice containing  $I_j$ . Diagrammatically, we can go:



That is, if  $m \in I_k$ , then  $n_1 + m + n_2 \in I_j$ , or  $I_k + n_1 + n_2 \subset I_j$ . Hence  $L_{d_1} \subset L_{d_2}$ , and  $d_2 \geq d_1$ . The converse argument gives  $d_1 \geq d_2$ , so  $d_1 = d_2$ . Q.E.D.

Let  $J$  be the set of recurrent states. Communication ( $\leftrightarrow$ ) is clearly an equivalence relationship on  $J$ , so splits  $J$  into disjoint classes  $C_1, C_2, \dots$ .

**Proposition 7.31.** For any  $C_i$

$$\sum_{k \in C_i} p(k | j) = 1, \quad \text{all } j \in C_i,$$

that is, each  $C_i$  is a closed set of states.

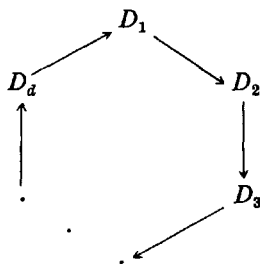
*Proof.* If  $j$  is recurrent and  $j \rightarrow k$ , then  $k \rightarrow j$ . Otherwise,  $P_j(X_n = j \text{ i.o.}) < 1$ . Hence the set of states  $\{k; j \rightarrow k\} = \{k; j \leftrightarrow k\}$ , but this latter is exactly the equivalence class containing  $j$ . The sum of  $p(k | j)$  over all  $k$  such that  $j \rightarrow k$  is clearly one.

Take  $C$  to be a closed indecomposable set of recurrent states. They all have the same period  $d$ . If  $d > 1$ , define the relationship  $\overset{d}{\leftrightarrow}$  as  $j \overset{d}{\leftrightarrow} k$  if  $p^{(n_1 d)}(k | j) > 0$ ,  $p^{(n_2 d)}(j | k) > 0$  for some  $n_1, n_2$ . Since  $j \overset{d}{\leftrightarrow} k$  is an equivalence relationship,  $C$  may be decomposed into disjoint sets  $D_1, D_2, \dots$  under  $\overset{d}{\leftrightarrow}$ .

**Proposition 7.32.** There are  $d$  disjoint equivalence classes  $D_1, D_2, \dots, D_d$  under  $\overset{d}{\leftrightarrow}$  and they can be numbered such that

$$\sum_{k \in D_{i+1} [d]} p(k | j) = 1, \quad j \in D_i$$

or diagrammatically,



The  $D_1, \dots, D_d$  are called cyclically moving subsets.

*Proof.* Denote by  $j \xrightarrow{[d]} k$  the existence of an  $n$  such that  $n[d] = [d]$  and  $p^{(n)}(k | j) > 0$ . Fix  $j_1$  and number  $D_1$  as the equivalence class containing  $j_1$ . Take  $j_2$  such that  $p(j_2 | j_1) > 0$ , and number  $D_2$  the equivalence class containing  $j_2$ , and so on. See that  $j \xrightarrow{0[d]} k$  implies that  $k \xrightarrow{0[d]} j$ . But

$$p^d(j_{d+1} | j_1) \geq p(j_{d+1} | j_d) \cdots p(j_2 | j_1) > 0,$$

so  $j_1 \xleftrightarrow{d} j_{d+1} \Rightarrow D_{d+1} = D_1$ . Consider any state  $k$ , such that  $p(k | j) > 0$ ,  $j \in D_1$  and  $k \notin D_2$ , say  $k \in D_3$ . Then  $k \xleftrightarrow{d} j_3$ . Look at the string

$$j_1 \xrightarrow{0[d]} j \xrightarrow{1[d]} k \xrightarrow{0[d]} j_3 \xrightarrow{1[d]} j_4 \rightarrow \cdots \xrightarrow{1[d]} j_{d+1} \xrightarrow{0[d]} j_1.$$

This string leads to  $j_1 \xrightarrow{d-1[d]} j_1$ . From this contradiction, conclude  $k \in D_2$ , and  $\sum_{k \in D_2} p(k | j) = 1, j \in D_1$ .

If  $C$  is a closed set of communicating nonperiodic states, one useful result is that for any two states  $j, k$ , all other states are common descendants. That is:

**Proposition 7.33.** *For  $j, k, l$ , any three states, there exists an  $n$  such that  $p^{(n)}(l | j) > 0, p^{(n)}(l | k) > 0$ .*

*Proof.* Take  $n_1, n_2$  such that  $p^{(n_1)}(l | j) > 0, p^{(n_2)}(l | k) > 0$ . Consider the set  $J = \{n; p^{(n)}(l | l) > 0\}$ . By the nonperiodicity, the smallest lattice containing  $J$  is  $L_1$ . Under addition,  $J$  is closed, so every integer can be expressed as  $s_1 m_1 - s_2 m_2$ ,  $m_1, m_2 \in J$ ,  $s_1, s_2$  nonnegative integers. Take  $n_2 - n_1 = s_1 m_1 - s_2 m_2$ , so  $n_1 + s_1 m_1 = n_2 + s_2 m_2 = n$ . Now check that  $p^{(n)}(l | j) > 0, p^{(n)}(l | k) > 0$ .

## Problems

8. If the set of states is finite, prove that

- there must be at least one recurrent state;
- every recurrent state is positive recurrent;
- there is a random variable  $n^*$  such that for  $n \geq n^*$ ,  $X_n$  is in the set of recurrent states.

9. Give an example of a chain with all states transient.

## 7. STATIONARY INITIAL DISTRIBUTIONS

Consider a chain  $X_0, X_1, \dots$  starting from an initial state  $i$  such that  $i$  is recurrent. Let  $N_n(j)$  be the number of visits of  $X_1, \dots, X_n$  to  $j$ , and  $\pi(j)$  be the expected number of visits to  $j$  before return to  $i$ ,  $\pi(i) = 1$ . The proof of Theorem 3.45 goes through, word for word. Use (3.46) again to conclude

$$\frac{N_n(j)}{N_n(l)} \xrightarrow{\text{a.s.}} \frac{\pi(j)}{\pi(l)}.$$

The relevance here is

**Proposition 7.34**

$$\pi(k) = \sum_j p(k | j) \pi(j)$$

for all  $k$  such that  $i \rightarrow k$ .

*Proof.* A visit to state  $j$  occurs on the  $n$ th trial before return to state  $i$  if  $\{X_n = j, X_{n+1} \neq i, \dots, X_1 \neq i\}$ . Therefore

$$\pi(j) = \sum_{n=1}^{\infty} P_i(X_n = j, X_{n-1} \neq i, \dots, X_1 \neq i),$$

so that

$$\begin{aligned} \sum_j p(k | j) \pi(j) &= \sum_{n=1}^{\infty} \sum_{j \neq i} P(X_{n+1} = k | X_n = j) P_i(X_n = j, X_{n-1} \neq i, \dots, X_1 \neq i) \\ &\quad + p(k | i) \\ &= \sum_{n=1}^{\infty} P_i(X_{n+1} = k, X_n \neq i, X_{n-1} \neq i, \dots, X_1 \neq i) \\ &\quad + P_i(X_1 = k) \\ &= \pi(k), \quad k \neq i. \end{aligned}$$

For  $k = i$ ,  $P_i(X_{n+1} = i, X_n \neq i, \dots, X_1 \neq i) = P_i(T^{(i)} = n + 1)$ , the right-hand side becomes  $\sum_{n=1}^{\infty} P_i(T^{(i)} = n) = 1$ , where  $T^{(i)}$  is the time of first recurrence of state  $i$ .

The  $\{\pi(j)\}$ , therefore, form a stationary initial *measure* for the chain. By summing, we get

$$\sum_{j \neq i} \pi(j) = \sum_{n=1}^{\infty} P_i(X_n \neq i, \dots, X_1 \neq i) = \sum_{n=1}^{\infty} P_i(T^{(i)} > n) = E_i T^{(i)} - 1$$

or

$$\sum_j \pi(j) = E_i T^{(i)}.$$

If  $i$  is positive-recurrent, then  $\bar{\pi}(j) = \pi(j)/E_i T^{(i)}$  forms a stationary initial distribution for the process. By Proposition 6.38, if  $T^{(j)}$  is the first recurrence time for state  $j$ , starting from  $j$ , then

$$\bar{\pi}(j) = \frac{1}{E_j T^{(j)}}.$$

Every equivalence class  $C$  of positive-recurrent states thus has the unique stationary initial distribution given by  $\bar{\pi}(j)$ . Note that  $\bar{\pi}(j) > 0$  for all  $j$  in

the class. Use the ergodic theorem to conclude

$$\frac{1}{n} \sum_{m=1}^n p^{(m)}(j | k) \rightarrow \bar{\pi}(j), \quad \text{all } k \in C.$$

Restrict the state space to such a class  $C$ .

**Proposition 7.35.** Let  $\lambda(j)$  be a solution of

$$\lambda(j) = \sum_k p(j | k) \lambda(k)$$

such that  $\sum_j |\lambda(j)| < \infty$ . Then there is a constant  $c$  such that  $\lambda(j) = c\bar{\pi}(j)$ .

*Proof.* By iteration

$$\lambda(j) = \sum_k p^m(j | k) \lambda(k), \quad m \geq 1.$$

Consequently,

$$\lambda(j) = \sum_k \left( \frac{1}{n} \sum_{m=1}^n p^{(m)}(j | k) \right) \lambda(k).$$

The inner term converges to  $\bar{\pi}(j)$ , and is always  $\leq 1$ . Use the bounded convergence theorem to get

$$\lambda(j) = \bar{\pi}(j) \sum_k \lambda(k).$$

This proposition is useful in that it permits us to get the  $\bar{\pi}(j)$  by solving the system  $\bar{\pi}(j) = \sum_k p(j | k) \bar{\pi}(k)$ .

## 8. SOME EXAMPLES

**Example A. Simple symmetric random walks.** Denote by  $I$  the integers, and by  $I^{(m)}$  the space  $\mathbf{j} = (j_1, \dots, j_m)$  of all  $m$ -tuples of integers. If the particle is at  $\mathbf{j}$ , then it makes the transition to any one of the  $2m$  nearest neighbors  $(j_1 \pm 1, j_2, \dots, j_m)$ ,  $(j_1, j_2 \pm 1, \dots, j_m), \dots$  with probability  $1/2m$ . The distribution of this process starting from  $\mathbf{j}$  is given by  $\mathbf{j} + \mathbf{Y}_1 + \dots + \mathbf{Y}_n$ , where  $\mathbf{Y}_1, \mathbf{Y}_2, \dots$  are independent identically distributed random vectors taking values in  $(\pm 1, 0, \dots, 0)$ ,  $(0, \pm 1, 0, \dots, 0)$  with equal probabilities. All states communicate. The chain has period  $d = 2$ . Denote  $\mathbf{0} = (0, 0, \dots, 0)$ . For  $m = 1$ , the process starting from  $\mathbf{0}$  is the fair coin-tossing game. Thus for  $m = 1$ , all states are null-recurrent. Polya [116] discovered the interesting phenomenon that if  $m = 2$ , all states are again null-recurrent, but for  $m \geq 3$  all states are transient. In fact, every random walk on  $I^{(m)}$  that is genuinely  $m$ -dimensional is null-recurrent for  $m \leq 2$ , transient for  $m \geq 3$ . See Chung and Fuchs [18].

**Example B. The renewal chain.** Another way of looking at a renewal process which illuminates the use of the word *renewal* is the following: At time zero, a new light bulb is placed in a fixture. Let  $T_1$  be the number of periods (integral) that it lasts. When it blows out at time  $n$ , it is replaced by another light bulb starting at time  $n$  that lasts  $T_2$  periods; the  $k$ th light bulb lasts  $T_k$  periods. The light bulbs are of identical manufacture and a reasonable model is to assume that the  $T_1, T_2, \dots$  are independent and identically distributed random variables. Also assume each bulb lasts at least one period; that is,  $P(T_1 \geq 1) = 1$ . Let  $A_n$  be the event that a light bulb blows out at time  $n$ . Intuitively, this starts the whole process over again. Mathematically, it is easy to show that  $\{A_n\}$  form a renewal event. Formally,  $A_n = \{\omega; \exists \text{ a } k \text{ such that } T_1 + \dots + T_k = n\}$ . Now the point is that 7.25 shows that the converse is true; given any renewal process  $U_1, U_2, \dots$ , and letting  $T_1, T_2, \dots$  be the times between occurrences of  $\{U_k = 1\}$ , we find that if  $P(T_1 < \infty) = 1$ , then

$$\{U_n = 1\} = \{\omega; \exists \text{ a } k \text{ such that } T_1 + \dots + T_k = n\},$$

where the  $T_1, T_2, \dots$  are independent and identically distributed.

For a Markov chain  $X_0, X_1, \dots$  starting from state  $j$ , the events  $\{X_n = j\}$  form a renewal event. Now we ask the converse question, given a renewal process  $U_1, U_2, \dots$ , is there a Markov process  $X_0, X_1, \dots$  starting from the origin such that the process  $\hat{U}_1, \hat{U}_2, \dots$ , defined by

$$\hat{U}_n = \begin{cases} 1, & X_n = 0, \\ 0, & \text{otherwise,} \end{cases}$$

has the same distribution as  $U_1, U_2, \dots$ ? Actually, we can define a Markov process  $X_0, X_1, \dots$  on the same sample space as the renewal process such that

$$\{U_n = 1\} = \{X_n = 0\}.$$

**Definition 7.36.** For a renewal process  $U_1, U_2, \dots$ , add the convention  $R_0 \equiv 0$ , and define the time of the last replacement prior to time  $n$  as

$$\tau_n = R_k \text{ on the set } \{\omega; R_k \leq n < R_{k+1}\}.$$

The age of the current item is defined by

$$X_n = n - \tau_n.$$

Clearly,

$$\{X_n = 0\} = \{\tau_n = n\} = \{\exists \text{ a } k \text{ such that } R_k = n\} = \{U_n = 1\}.$$

The  $X_n$  process, as defined, takes values in the nonnegative integers, and  $X_0 \equiv 0$ . What does it look like? If  $X_n = j$ , then  $\tau_n = n - j$  and on this set there exists no  $k$  such that  $n - j < R_k \leq n$ . Therefore, on the set  $\tau_n = n - j$ , either  $\tau_{n+1} = n - j$  or  $\tau_{n+1} = n + 1$ . So if  $X_n = j$ , either

$X_{n+1} = j + 1$  or  $X_{n+1} = 0$ . Intuitively, either a renewal takes place at time  $n + 1$  or the item ages one more time unit. Clearly,  $X_0, X_1, \dots$  is a Markov chain with the stationary transition probabilities

$$\begin{aligned} P(X_{n+1} = j + 1 | X_n = j) &= P(\tau_{n+1} = n - j | \tau_n = n - j) \\ &= P(T_1 > j + 1 | T_1 > j) \\ &= \frac{P(T_1 > j + 1)}{P(T_1 > j)}. \end{aligned}$$

All states communicate, the chain has period determined by the minimum lattice on which  $T_1$  is distributed, and is null-recurrent or positive-recurrent as  $ET_1 = \infty$  or  $< \infty$ .

If  $ET_1 < \infty$ , there is a stationary point process  $U_1^*, U_2^*, \dots$  having interpoint distances  $T_1^*, T_2^*, \dots$  with the same distribution as  $T_1, T_2, \dots$ . For this process use  $R_k^*$  as the time of the  $k$ th point past the origin, and define  $\tau_n^*, X_n^*$  as above. The transition probabilities of  $X_n^*$  are the same as for  $X_n$ , but  $X_n^*$  is easily seen to be stationary. At  $n = 0$ , the age of the current item is  $k$  on the set  $\{U_0^* = 0, \dots, U_{-k+1}^* = 0, U_{-k}^* = 1\}$ . The probability of this set is, by stationarity,

$$P^*(U_k^* = 0, \dots, U_1^* = 0, U_0^* = 1) = P(T_1 > k)/ET_1.$$

Hence  $\pi(k) = P(T_1 > k)/ET_1$  is a stationary initial distribution for the process. The question of asymptotic stationarity of  $X_0, X_1, \dots$  is equivalent to asking if the renewal process is asymptotically stationary in the sense

$$P((U_n, U_{n+1}, \dots) \in B) \rightarrow P^*((U_1^*, \dots) \in B)$$

for every  $B \in \mathcal{B}_\infty(\{0, 1\})$ .

**Problem 10.** Show that a sufficient condition for the  $U_1, U_2, \dots$  process to be asymptotically stationary in the above sense is

$$P(U_n = 1) \rightarrow \frac{1}{ET_1}.$$

**Example C. Birth and death processes.** These are a class of Markov chains in which the state space is the integers  $I$  or the nonnegative integers  $I^+$  and where, if the particle is at  $j$ , it can move either to  $j + 1$  with probability  $\alpha_j$ , to  $j - 1$  with probability  $\beta_j$ , or remain at  $j$  with probability  $1 - \alpha_j - \beta_j$ . If the states are  $I$ , assume all states communicate. If the states are  $I^+$ , 0 is either an *absorbing state* (defined as any state  $i$  such that  $p(i | i) = 1$ ) or *reflecting* ( $p(1 | 0) > 0$ ). Assume that all other states communicate between themselves, and can get to zero. Equivalently,  $\alpha_j \neq 0, \beta_j \neq 0$ , for  $j > 0$ . If 0 is absorbing, then all other states are transient, because  $j \rightarrow 0$  but  $0 \nrightarrow j, j \neq 0$ . Therefore, for almost every sample path, either  $X_n \rightarrow \infty$  or  $X_n \rightarrow 0$ . If 0 is reflecting, the states can be transient or recurrent, either positive or null.



To get a criterion, let  $\tau_j^*$  be the first passage time to state  $j$  starting from zero:

$$\tau_j^* = \min \{n; X_n = j\}, X_0 \equiv 0.$$

Let  $A_j$  be the event that a return to zero occurs between  $\tau_j^*$  and  $\tau_{j+1}^*$ ,

$$A_j = \bigcup_n \{X_n = 0, \tau_j^* \leq n \leq \tau_{j+1}^*\}.$$

The  $\tau_j^*$  are Markov times, and we use 7.8 to conclude that the  $A_j$  are independent events. By the Borel-Cantelli lemma, the process is transient or recurrent as  $\sum_1^\infty P(A_j) < \infty$  or  $\sum_1^\infty P(A_j) = \infty$ . Let  $\tau^*$  be the first time after  $\tau_j^*$  that  $X_n \neq j$ . Then  $P(A_j) = E(P(A_j | X_{\tau^*}))$ . Now

$$P(A_j | X_{\tau^*} = j+1) = 0.$$

On the set  $X_{\tau^*} = j-1$ ,  $A_j$  can occur if we return to zero before climbing to  $j$  or by returning to zero only after climbing to  $j$  but before climbing to  $j+1$ . Since  $\tau^*$  is a Markov time, by the strong Markov property

$$P(A_j | X_{\tau^*} = j-1) = P(A_{j-1}) + P(A_{j-1}^c)P(A_j), \quad j > 1$$

Checking that  $P(X_{\tau^*} = j-1) = \beta_j/(\alpha_j + \beta_j)$  gives the equation

$$P(A_j) = \frac{\beta_j}{\alpha_j + \beta_j} (P(A_{j-1}) + P(A_{j-1}^c)P(A_j)),$$

or

$$P(A_j) = \frac{r_j P(A_{j-1})}{1 + r_j P(A_{j-1})}$$

where  $r_j = \beta_j/\alpha_j$ . Direct substitution verifies that

$$P(A_j) = \frac{\rho_j}{s_j}; \quad \text{where } \rho_j = \prod_1^j \frac{\beta_k}{\alpha_k}, \quad s_j = 1 + \sum_1^j \rho_k.$$

Certainly, if  $\sum_1^\infty \rho_j < \infty$  then  $\sum_1^\infty P(A_j) < \infty$ . To go the other way, note that since  $s_j = s_{j-1}/(1 - \rho_j/s_j)$ , then

$$\sum_1^\infty \frac{\rho_j}{s_j} < \infty \Rightarrow \sum_2^\infty \frac{\rho_j}{s_{j-1}} < \infty.$$

But

$$\sum_2^k \frac{\rho_j}{s_{j-1}} \geq \sum_2^k \log(1 + \frac{\rho_j}{s_{j-1}}) = \log(s_k) - \log(s_1).$$

We have proved

**Proposition 7.37.** *A birth and death process on  $I^+$  with the origin reflecting is transient iff*

$$\sum_{j=1}^\infty \prod_{k=1}^j \left( \frac{\beta_k}{\alpha_k} \right) < \infty.$$

(Due to Harris [68]. See Karlin [86, p. 204] for an alternative derivation.) To discriminate between null and positive recurrence is easier.

**Problem 11.** Use the condition that

$$\lambda(k) = \sum_j p(k|j) \lambda(j)$$

has no solutions such that  $\sum |\lambda(k)| < \infty$  to find a necessary and sufficient condition that a recurrent birth and death process on  $I^+$  be null-recurrent.

**Example D. Branching processes.** These processes are characterized as follows: If at time  $n$  there are  $k$  individuals present, then the  $j$ th one independently of the others gives birth to  $Y_j$  offspring by time  $n + 1$ ,  $j = 1, \dots, k$ , where  $P(Y_j = l) = p_l$ ,  $l = 0, 1, \dots$ . The  $\{Y_j = 0\}$  event corresponds to the death of the  $j$ th individual leaving no offspring. The state space is  $I^+$ , the transition probabilities for  $X_n$ , the population size at time  $n$ , are

$$p(m | k) = P(Y_1 + \dots + Y_k = m),$$

where the  $Y_1, \dots, Y_k$  are independent and have the same distribution as  $Y_1$ . Zero is an absorbing state (unless the model is revised to allow the introduction of new individuals into the population). If  $p_0 > 0$ , then the same argument as for birth and death processes establishes the fact that every state except zero is transient. If  $p_0 = 0$ , then obviously the same result holds. For a complete and interesting treatment of these chains and their generalizations, see Harris [67].

**Problem 12.** In a branching process, suppose  $EY_1 = m < \infty$ . Use the martingale convergence theorem to show that  $X_n/m^n$  converges a.s.

**Example E. The Ehrenfest urn scheme.** Following the work of Gibbs and Boltzmann statistical mechanics was faced with this paradox. For a system of particles in a closed container, referring to the  $6N$  position-velocity vector as the state of the system, then in the ergodic case every state is recurrent in the sense that the system returns infinitely often to every neighborhood of any initial state.

On the other hand, the observed macroscopic behavior is that a system seems to move irreversibly toward an equilibrium condition. Smoluchowski proposed the solution that states far removed from equilibrium have an enormously large recurrence time, thus the system over any reasonable observation time appears to move toward equilibrium. To illustrate this the Ehrenfests constructed a model as follows: consider two urns I and II, and a total of  $2N$  molecules distributed within the two urns. At time  $n$ , a molecule is chosen at random from among the  $2N$  and is transferred from whatever urn it happens to be in to the other urn. Let the state  $k$  of the chain be the number of molecules in urn I,  $k = 0, \dots, 2N$ . The transition probabilities are given by

$$p(k+1 | k) = \frac{2N-k}{2N}, \quad p(k-1 | k) = \frac{k}{2N}.$$

All states communicate, and since there are only a finite number, all are positive-recurrent. We can use the fact that the stationary distribution  $\bar{\pi}(k) = 1/E_k T^{(k)}$  to get the expected recurrence times.

**Problem 13.** Use the facts that

$$\bar{\pi}(k) = \sum_j p(k | j) \bar{\pi}(j), \quad \sum_k \bar{\pi}(k) = 1$$

to show that

$$a) \quad \bar{\pi}(N+k) = \frac{(2N)!}{(N+k)!(N-k)!} \cdot 2^{-2N}.$$

Compare this with the derivation of the central limit theorem for coin-tossing, Chapter I, Section 3, and show that for  $N$  large, if  $T$  is the recurrence time for the states  $\{k; |N - k| > x\sqrt{N/2}\}$

$$b) \quad ET \simeq \frac{1}{\sqrt{2/\pi} \int_x^\infty e^{-t^2/2} dt}.$$

See Kac [80] for further discussion.

## 9. THE CONVERGENCE THEOREM

The fundamental convergence result for Markov chains on the integers is

**Theorem 7.38.** *Let  $C$  be a closed indecomposable set of nonperiodic recurrent states. If the states are null-recurrent, then for all  $j, k \in C$*

$$p^{(n)}(j | k) \rightarrow 0.$$

*If the states are positive-recurrent with stationary initial distribution  $\bar{\pi}(j)$ , then for all  $j, k \in C$*

$$p^{(n)}(j | k) \rightarrow \bar{\pi}(j).$$

There are many different proofs of this. Interestingly enough, the various proofs are very diverse in their origin and approach. One simple proof is based on

**Theorem 7.39** (The renewal theorem). *For a nonperiodic renewal process,*

$$P(U_n = 1) \rightarrow \begin{cases} \frac{1}{ET_1}, & \text{if } ET_1 < \infty, \\ 0, & \text{if } ET_1 = \infty. \end{cases}$$

There is a nice elementary proof of this in Feller [59, Volume I], and we prove a much generalized version in Chapter 10.

The way this theorem is used in 7.38 is that for  $\{U_n\}$  the return process for state  $j$ ,  $P_j(U_n = 1) = p^{(n)}(j | j)$ ; hence  $p^{(n)}(j | j) \rightarrow \bar{\pi}(j)$  if  $E_j T^{(j)} < \infty$ , or  $p^{(n)}(j | j) \rightarrow 0$  if  $j$  is null-recurrent. No matter where the process is started, let  $T^{(j)}$  be the first time that  $j$  is entered. Then

$$(7.40) \quad p^{(n)}(j | k) = \sum_{m=1}^n p^{(n-m)}(j | j) P_k(T^{(j)} = m)$$

by the Markov property. Argue that  $P_k(T^{(j)} < \infty) = 1$  (see the proof of 7.30). Now use the bounded convergence theorem to establish

$$\lim_n p^{(n)}(j | k) = \lim_n p^{(n)}(j | j) \sum_{m=1}^{\infty} P_k(T^{(j)} = m).$$

We can also use some heavier machinery to get a stronger result due to Orey [114]. We give this proof, from Blackwell and Freedman [9], because it involves an interesting application of martingales and the Hewitt-Savage zero-one law.

**Theorem 7.41.** *Let  $C$  be a closed indecomposable set of nonperiodic recurrent states. Then for any states  $k, l$  in  $C$*

$$\sum_{j \in C} |p^{(n)}(j | k) - p^{(n)}(j | l)| \rightarrow 0.$$

**Remark.** We can get 7.38 in the positive-recurrent case from 7.41 by noting that 7.41 implies

$$\sum_{i \in C} (p^{(n)}(j | k) - p^{(n)}(j | l)) \bar{\pi}(l) \rightarrow 0.$$

In the null-recurrent case we use an additional fact: First, consider the event  $A_m$  that starting from  $j$  the last entry into  $j$  up to time  $n$  was at time  $n - m$ ,  $A_m = \{X_{n-m} = j, X_{n-m+1} \neq j, \dots, X_n \neq j\}$ . The  $A_m$  are disjoint and the union is the whole space. Furthermore, for the process starting from  $j$ ,

$$\begin{aligned} P_j(A_m) &= P_j(X_{n-m} = j) P_j(X_{n-m+1} \neq j, \dots, X_n \neq j | X_{n-m} = j) \\ &= p^{(n-m)}(j | j) P_j(T^{(j)} > m). \end{aligned}$$

Consequently,

$$(7.42) \quad 1 = \sum_{m=0}^n p^{(n-m)}(j | j) P_j(T^{(j)} > m),$$

(where  $p^{(0)}(j | j) = 1$ ).

Let  $\limsup p^{(n)}(j | j) = \bar{p}$ . Take a subsequence  $n'$  such that  $p^{(n')}(j | j) \rightarrow \bar{p}$ . By 7.41, for any other state,  $p^{(n')}(j | k) \rightarrow \bar{p}$ . Use this to get

$$p^{(n'+1)}(j | j) = \sum_k p^{(n')}(j | k) p(k | j) \rightarrow \bar{p}.$$

Then for any  $r \geq 0$  fixed,  $p^{(n'+r)}(j | j) \rightarrow \bar{p}$ . Substitute  $n = n' + r$  in (7.42), and chop off some terms to get

$$1 \geq \sum_{m=0}^r p^{(n'+r-m)}(j | j) P_j(T^{(j)} > m) \rightarrow \bar{p} \sum_{m=0}^r P_j(T^{(j)} > m).$$

Noting that

$$\sum_0^\infty P_j(T^{(j)} > m) = E_j T^{(j)} = \infty$$

implies  $\bar{p} = 0$ , and using 7.41 we can complete the proof that  $p^{(n)}(j | k) \rightarrow 0$ .

*Proof of 7.41.* Take  $\pi$  any initial distribution such that  $\pi(l) > 0$ . Let  $\mathcal{F}$  be the tail  $\sigma$ -field of the process  $X_0, X_1, \dots$  and suppose that  $\mathcal{F}$  has the zero-one property under  $P_\pi$ . Then an easy consequence of the martingale convergence theorem (see Problem 6, Chapter 5) is that for any  $A \in \mathcal{F}(X)$ ,

$$\sup_{B \in \mathcal{F}(X_n, \dots)} |P_\pi(A \cap B) - P_\pi(A)P_\pi(B)| \rightarrow 0.$$

Let  $D \subset C$ ,  $B = \{X_n \in D\}$ ,  $A = \{X_0 = l\}$ , so that

$$\sup_{D \subset C} |P_\pi(X_n \in D, X_0 = l) - P_\pi(X_n \in D)\pi(l)| \rightarrow 0$$

or

$$\sup_{D \subset C} |P_l(X_n \in D) - P_\pi(X_n \in D)| \rightarrow 0.$$

Write  $C_n^+ = \{j; P_l(X_n = j) \geq P_\pi(X_n = j)\}$ ,  $C_n^-$  the complement of  $C_n^+$  in  $C$ . Then by the above,

$$|P_l(X_n \in C_n^+) - P_\pi(X_n \in C_n^+)| \rightarrow 0, \quad |P_l(X_n \in C_n^-) - P_\pi(X_n \in C_n^-)| \rightarrow 0,$$

implying

$$\sum_{j \in C} |P_l(X_n = j) - P_\pi(X_n = j)| \rightarrow 0.$$

Now use the initial distribution  $\pi$  which assigns mass  $\frac{1}{2}$  each to the states  $l$  and  $k$  to get the stated result.

The completed proof is provided by

**Theorem 7.43.** *For any tail event  $A$ , either  $P_j(A)$  is one for all  $j \in C$ , or zero for all  $j \in C$  (under the conditions of 7.41).*

*Proof.* Consider the process starting from  $j$ . The random vectors  $Z_k = (X_{R_k}, \dots, X_{R_{k+1}-1})$  are independent and identically distributed by 7.28. Clearly,  $\mathcal{F}(X) = \mathcal{F}(Z_0, Z_1, \dots)$ . Take  $W$  a tail random variable; that is,  $W$  is measurable  $\mathcal{F}$ .

For every  $n$ , there is a random variable  $\varphi_n(x)$  on  $(R^{(\infty)}, \mathcal{B}_\infty)$  such that  $W = \varphi_n(X_n, \dots)$ . So for every  $k$ ,

$$W = \varphi_{R_k}(X_{R_k}, X_{R_{k+1}}, \dots) = \theta(R_k, Z_k, Z_{k+1}, \dots).$$

Now  $R_k$  is a symmetric function of  $Z_0, \dots, Z_{k-1}$ . Hence  $W$  is a symmetric function of  $Z_0, Z_1, \dots$ . The Hewitt-Savage zero-one law holds word by word for independent identically distributed random vectors, instead of random variables. Therefore  $W$  is a.s. constant, and  $P_j(A)$  is zero or one for every tail event  $A$ .

For any two states  $j, k$  let  $l$  be a descendent,  $p^{(n)}(l | j) > 0, p^{(n)}(l | k) > 0$ . Write

$$P_j(A) = E_j(P_j(A | X_n, \dots, X_0)).$$

Since  $A$  is measurable  $\mathcal{F}(X_{n+1}, \dots)$ ,

$$P_j(A) = E_j(P_j(A | X_n)) = \sum_i P_j(A | X_n = i) p^{(n)}(i | j).$$

If  $P_j(A) = 1$ , then  $P_j(A | X_n = l) \neq 0$ . But using  $k$  instead of  $j$  above, we get

$$P_k(A) \geq P_k(A | X_n = l) p^{(n)}(l | k) > 0.$$

Hence  $P_k(A) = 1$ , and the theorem is proved.

From this fundamental theorem follows a complete description of the asymptotic behavior of the  $p^{(n)}(j | k)$ . If a closed communicating set of positive recurrent states has period  $d$ , then any one of the cyclically moving subclasses  $D_r$ ,  $r = 1, \dots, d$  is nonperiodic and closed under the transition probability  $p^{(d)}(j | k)$ . Looking at this class at time steps  $d$  units apart, conclude that

$$p^{(nd)}(j | k) \rightarrow \frac{d}{E_j T^{(j)}}, \quad j, k \in D_r.$$

If  $k \xrightarrow{[d]} j$ , use (7.40) to get

$$p^{(nd+1)}(j | k) \rightarrow \frac{d}{E_j T^{(j)}}.$$

If both transient states and positive-recurrent states are present then the asymptotic behavior of  $p^{(n)}(j | k)$ ,  $j$  positive-recurrent and nonperiodic,  $k$  transient, depends on the probability  $P(C | k)$  that starting from  $k$  the process will eventually enter the class of states  $C$  communicating with  $j$ . From (7.40), in fact,

$$p^{(n)}(j | k) \rightarrow \frac{P(C | k)}{E_j T^{(j)}}.$$

When  $j$  is periodic, the behavior depends not only on  $P(C | k)$  but also at what point in the cycle of motion in  $C$  the particle from  $k$  enters  $C$ .

## 10. THE BACKWARD METHOD

There is a simple device which turns out to be important, both theoretically and practically, in the study of Markov chains. Let  $Z = \varphi(X_0, X_1, \dots)$  be any random variable on a Markov chain with stationary transition probabilities. Then the device is to get an equation for  $f(x) = E(Z | X_0 = x)$  by using the fact that

$$E(E(Z | X_1 = y, X_0 = x) | X_0 = x) = E(Z | X_0 = x).$$

Of course, this will be useful only if  $E(Z | X_1 = y, X_0 = x)$  can be expressed in terms of  $f$ . The reason I call this the backward method is that it is the initial conditions of the process that are perturbed. Here are some examples.

Mostly for convenience, look at the countable cases. It is not difficult to see how the same method carries over to similar examples in the general case.

a) *Invariant random variable.* Let  $Z$  be a bounded invariant function,  $Z = \varphi(X_n, \dots)$ ,  $n \geq 0$ . Then if  $E(Z | X_0 = j) = f(j)$ ,

$$E(Z | X_1 = k, X_0 = j) = E(\varphi(X_1, \dots) | X_1 = k) = f(k),$$

so that  $f(j)$  is a bounded solution of

$$(7.44) \quad f(j) = \sum_k f(k)p(k | j).$$

There is an interesting converse. Let  $f(j)$  be a bounded solution of (7.44). Then write (7.44) as

$$f(X_n) = E(f(X_{n+1}) | X_n).$$

By the Markov property,

$$f(X_n) = E(f(X_{n+1}) | X_n, \dots, X_0).$$

This says that  $f(X_n)$  is a martingale. Since it is bounded, the convergence theorem applies, and there is a random variable  $Y$  such that

$$f(X_n) \xrightarrow{\text{a.s.}} Y$$

If  $Y = \theta(X_0, X_1, \dots)$ , from

$$f(X_{n+1}) \xrightarrow{\text{a.s.}} Y$$

conclude that  $Y = \theta(X_1, X_2, \dots)$  a.s. Thus  $Y$  is a.s. invariant, and

$$f(j) = E(Y | X_0 = j).$$

(Use here an initial distribution which assigns positive mass to all states.) Formally,

**Proposition 7.45.** *Let  $\pi(j) > 0$  for all  $j \in F$ , then there is a one-to-one correspondence between bounded a.s. invariant random variables and bounded solutions of (7.44).*

b) *Absorption probabilities.* Let  $C_1, C_2, \dots$  be closed sets of communicating recurrent states. Let  $A$  be the event that a particle starting from state  $k$  is eventually absorbed in  $C_1$ ,

$$A = \{X_n \in C_1, \text{ all } n \text{ sufficiently large}\}.$$

$A$  is an invariant event, so  $f(j) = P(A | X_0 = j)$  satisfies (7.44). There are also the boundary conditions:

$$f(j) = \begin{cases} 1, & j \in C_1, \\ 0, & j \in C_h, \quad h \neq 1. \end{cases}$$

If one solves (7.44) subject to these boundary conditions and boundedness, is the solution unique? No, in general, because if  $J$  is the set of all transient states, the event that the particle remains in  $J$  for all time is invariant, and

$$c(j) = P_j(X_n \in J, n = 0, 1, \dots)$$

is zero on all  $C_h$ , and any multiple of  $c(j)$  may be added to any given solution satisfying the boundary conditions to give another solution. If the probability of remaining in transient states for all time is zero, then the solution is unique. For example, let  $g$  be such a solution, and start the process from state  $j$ . The process  $g(X_n)$ ,  $n = 0, 1, \dots$  is a martingale. Let  $n^*$  be the first time that one of the  $C_h$  is entered. This means that  $n^*$  is a stopping time for the  $X_0, X_1, \dots$  sequence. Furthermore  $g(X_n)$  and  $n^*$  satisfy the hypothesis of the optional stopping theorem. Therefore,

$$Eg(X_{n^*}) = Eg(X_0) = g(j).$$

But

$$g(X_{n^*}) = \begin{cases} 1, & X_{n^*} \in C_i, \\ 0, & X_{n^*} \notin C_i. \end{cases}$$

Therefore  $g(j) = P(A | X_0 = j)$ .

In fair coin-tossing, with initial fortune zero, what is the probability that we win  $M_1$  dollars before losing  $M_2$ ? This is the same problem as: For a simple symmetric random walk starting from zero, with absorbing states at  $M_1, -M_2$ , find the probability of being absorbed into  $M_1$ . Let  $p^+(j)$  be the probability of being absorbed into  $M_1$  starting from  $j$ ,  $-M_2 \leq j \leq M_1$ . Then  $p^+(j)$  must satisfy (7.44) which in this case is

$$p^+(j) = \frac{1}{2}p^+(j-1) + \frac{1}{2}p^+(j+1), \quad -M_2 < j < M_1,$$

and the boundary conditions  $p^+(-M_2) = 0$ ,  $p^+(M_1) = 1$ . This solution is easy to get:

$$p^+(j) = \frac{j + M_2}{M_1 + M_2}.$$

c) *Two other examples.* Among many others, not involving invariant sets, I pick two. Let  $n^*$  be the time until absorption into the class  $C$  of recurrent states, assuming  $C \neq \emptyset$ . Write  $m(j) = E_j(n^*)$ . Check that

$$E(n^* | X_1 = k, X_0 = j) = \begin{cases} 1, & k \in C, \quad j \notin C, \\ 1 + m(k), & k \notin C, \quad j \notin C, \end{cases}$$

and apply the backward argument to give

$$(7.46) \quad m(j) = \sum_{k \notin C} m(k)p(k | j) + 1, \quad j \notin C.$$

The boundary conditions are  $m(j) = 0$ ,  $j \in C$ .



Now let  $N_i$  be the number of visits to state  $i$  before absorption into  $C$ . Denote  $G(j, i) = E_j(N_i)$ . For  $k \notin C, j \notin C$ ,

$$E_j(N_i | X_1 = k) = \begin{cases} G(k, i), & j \neq i, \\ G(k, i) + 1, & j = i. \end{cases}$$

So

$$(7.47) \quad G(j, i) = \delta(j, i) + \sum_{k \in C^c} G(k, i)p(k | j),$$

where  $\delta(j, i) = 0$  or  $1$  as  $j \neq i$  or  $j = i$ . The boundary conditions are  $G(j, i) = 0, j \in C$ . Of course, this makes no sense unless  $i$  is transient.

With these last two examples there is a more difficult uniqueness problem. For example, in (7.46) assume that

$$E_j(n^*) < \infty, \quad \text{all } j.$$

Then any nonnegative solution  $g(j)$  of (7.46) satisfying

$$\lim_{N \rightarrow \infty} \int_{\{n^* > N\}} g(X_N) dP = 0$$

must be  $E_j(n^*)$ . To prove this, check that

$$g(X_n) - \sum_{k=0}^{n-1} \chi_J(X_k)$$

is a martingale sequence, that  $E_j\left(\sum_{k=0}^{n^*-1} \chi_J(X_k)\right) = E_j(n^*)$ , and apply optional stopping.

## Problems

**14.** For simple symmetric random walk with absorbing states at  $M_1, -M_2$ , show that

$$\begin{aligned} \text{a)} \quad & E_j(n^*) = (M_2 + j)(M_1 - j), \\ \text{b)} \quad & E_j(N_i) = \begin{cases} \frac{(M_1 - j)(i + M_2)}{M_1 + M_2}, & j \geq i, \\ \frac{(M_2 + j)(M_1 - i)}{M_1 + M_2}, & j \leq i. \end{cases} \end{aligned}$$

**15.** Let  $\{X_n\}$  be simple symmetric random walk. Derive the expressions for  $p^+(j)$  and  $E_j(n^*)$  by showing that the sequences  $\{X_n\}, \{X_n^2 - n\}$  are martingales and applying the stopping time results of Section 7, Chapter 5.

**16.** For simple symmetric random walk with absorbing states at  $M_1, -M_2$ , use the expression for  $p^+(j)$  to evaluate  $q(j) = P_j$  (at least one return to  $j$ ). For  $-M_2 < j < i < M_1$ ,  $E_j(N_i)$  is the probability that particle hits  $i$

before  $-M_2$  times the expected number of returns to  $i$  starting from  $i$  before absorption. Use  $p^+(j)$ , for absorbing states at  $-M_2$ ,  $i$ , and  $q(i)$  to evaluate  $E_j(N_i)$ .

17. For any given set  $D$  of states, let  $A$  be the event that  $X_n$  stays in  $D$  for all  $n$ ,

$$A = \{X_n \in D, n = 0, 1, \dots\}.$$

a) Show that  $f(j) = P(A | j)$  satisfies

$$f(j) = \sum_{k \in D} f(k)p(k | j), \quad j \in D,$$

$$f(j) = 0, \quad j \in D^c.$$

b) Prove using (a) that a state  $h$  is transient iff there exists a bounded non-trivial solution to the equation

$$g(j) = \sum_{k \neq h} g(k)p(k | j), \quad j \neq h.$$

c) Can you use (b) to deduce 7.37?

## NOTES

In 1906 A. A. Markov [110] proved the existence of stationary initial distributions for Markov chains with a finite number of states. His method is simple and clever, and the idea can be generalized. A good exposition is in Doob's book [38, pp. 170 ff]. The most fundamental work on general state spaces is due to W. Doeblin [25] in 1937 and [28] in 1940. Some of these latter results concerning the existence of invariant initial distributions are given in Doob's book. The basic restriction needed is a sort of compactness assumption to keep the motion from being transient or null-recurrent. But a good deal of Doeblin's basic work occurs before this restriction is imposed, and is concerned with the general decomposition of the state space. For an exposition of this, see K. L. Chung [15] or [17].

The difficulty in the general state space is that there is no way of classifying each state  $y$  by means of the process of returns to  $y$ . If, for example,  $p(A | x)$  assigns zero mass to every one-point set, then the probability of a return to  $x$  is zero. You might hope to get around this by considering returns to a neighborhood of  $x$ , but then the important independence properties of the recurrence times no longer hold. It may be possible to generalize by taking smaller and smaller neighborhoods and getting limits, but this program looks difficult and has not been carried out successfully. Hence, in the general case, it is not yet clear what definition is most appropriate to use in classifying chains as recurrent or transient. For a fairly natural definition of recurrent chains Harris [66] generalized Doeblin's result by showing the existence of a

possibly infinite, but always  $\sigma$ -finite measure  $Q(dx)$  satisfying

$$Q(B) = \int p(B | x)Q(dx), \quad B \in \mathcal{B}_1(F).$$

His idea was very similar to the idea in the countable case: Select a set  $A \in \mathcal{B}_1(F)$  so that an initial distribution  $\pi_A(\cdot)$  exists concentrated on  $A$  such that every time the process returned to  $A$ , it had the distribution  $\pi_A$ . This could be done using Doeblin's technique. Then define  $\pi(B)$ ,  $B \in \mathcal{B}_1(F)$  as the expected number of visits to  $B$  between visits to  $A$ , using the initial distribution  $\pi_A$ .

The basic work when the state space is countable but not necessarily finite is due to Kolmogorov [95], 1936. The systematic application of the renewal theorem and concepts was done by Feller, see [55]. K. L. Chung's book [16] is an excellent source for a more complete treatment of the countable case.

The literature concerning applications of Markov chains is enormous. Karlin's book [86] has some nice examples; so does Feller's text [59, Vol. I]. A. T. Bharucha-Reid's book [3] is more comprehensive.

The proof of Proposition 7.37 given in the first edition of this book was incorrect. I am indebted to P. J. Thomson and K. M. Wilkinson for pointing out the error and supplying a correction.

# CHAPTER 8

## CONVERGENCE IN DISTRIBUTION AND THE TOOLS THEREOF

### 1. INTRODUCTION

Back in Chapter 1, we noted that if  $Z_n = Y_1 + \cdots + Y_n$ , where the  $Y_i$  are independent and  $\pm 1$  with probability  $\frac{1}{2}$ , then

$$P\left(\frac{Z_n}{\sqrt{n}} < x\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Thus the random variables  $Z_n/\sqrt{n}$  have distribution functions  $F_n(x)$  that converge for every value of  $x$  to  $\Phi(x)$ , but from Problem 23, Chapter 3, certainly the random variables  $Z_n/\sqrt{n}$  do not converge a.s. (or for that matter in  $L_1$ , or in any strong sense). What are convergent here are not the values of the random variables themselves, but the probabilities with which the random variables assume certain values. In general, we would like to say that the distribution of the random variable  $X_n$  converges to the distribution of  $X$  if  $F_n(x) = P(X_n < x) \rightarrow F(x) = P(X < x)$  for every  $x \in R^{(1)}$ . But this is a bit too strong. For instance, suppose  $X \equiv 0$ . Then we would want the values of  $X_k$  to be more and more concentrated about zero, that is for any  $\epsilon > 0$  we would want

$$(8.1) \quad P(X_n < -\epsilon) \rightarrow 0, \quad P(X_n \geq \epsilon) \rightarrow 0.$$

Now  $F(0) = 0$ , but 8.1 could hold, even with  $F_n(0) = 1$ , for all  $n$ . Take  $X_n \equiv -1/n$ , for example. What 8.1 says is that for all  $x < 0$ ,  $F_n(x) \rightarrow F(x)$ , and for all  $x > 0$ ,  $F_n(x) \rightarrow F(x)$ . Apparently, not much should be assumed about what happens for  $x$  a discontinuity point of  $F(x)$ . Hence we state the following:

**Definition 8.2.** We say that  $X_n$  converges to  $X$  in distribution,  $X_n \xrightarrow{D} X$ , if  $F_n(x) \rightarrow F(x)$  at every point  $x \in C(F)$ , the set of continuity points of  $F$ . That is,  $P(X = x) = 0 \Rightarrow F_n(x) \rightarrow F(x)$ . We will also write in this case  $F_n \xrightarrow{D} F$ .

Different terminology is sometimes used.

**Definition 8.3.** By the law of  $X$ , written  $\mathcal{L}(X)$ , is meant the distribution of  $X$ . Convergence in distribution is also called convergence in law and  $\mathcal{L}(X_n) \rightarrow \mathcal{L}(X)$  is equivalent notation for  $X_n \xrightarrow{D} X$ . If random variables  $X$  and  $Y$  have the same distribution, write either  $\mathcal{L}(X) = \mathcal{L}(Y)$  or  $X \stackrel{D}{=} Y$ .

Recall from Chapter 2 that a function  $F(x)$  on  $R^{(1)}$  is the distribution function of a random variable iff

- (8.4) i)  $x \leq y \Rightarrow F(x) \leq F(y)$   
 ii)  $x \uparrow y \Rightarrow F(x) \uparrow F(y)$   
 iii)  $F(-\infty) = 0, F(+\infty) = 1.$

### Problems

1. Show that if  $F(x) = P(X < x)$ , then  $F(x^+) - F(x) = P(X = x)$ . Show that  $C^c(F)$  is at most countable ( $F(x^+) = \lim_{y \downarrow x} F(y)$ ).
2. Let  $T$  be a dense set of points in  $R^{(1)}$ ,  $F_0(x)$  on  $T$  having properties (8.4 i, ii, and iii) with  $x, y \in T$ . Show that there is a unique distribution function  $F(x)$  on  $R^{(1)}$  such that  $F(x) = F_0(x)$ ,  $x \in T$ .
3. Show that if, for each  $n$ ,  $X_n$  takes values in the integers  $I$ , then  $X_n \xrightarrow{\mathcal{D}} X$  implies  $P(X \in I) = 1$  and  $X_n \xrightarrow{\mathcal{D}} X \Leftrightarrow P(X_n = j) \rightarrow P(X = j)$ , all  $j \in I$ .
4. If  $F_n, F$  are distribution functions,  $F_n \xrightarrow{\mathcal{D}} F$ , and  $F(x)$  continuous, show that

$$\sup_x |F_n(x) - F(x)| \rightarrow 0.$$

5. If  $X \stackrel{\mathcal{D}}{=} Y$ , and  $\varphi(x)$  is  $\mathcal{B}_1$ -measurable, show that  $\varphi(X) \stackrel{\mathcal{D}}{=} \varphi(Y)$ . Give an example to show that if  $X, Y, Z$  are random variables defined on the same probability space, that  $X \stackrel{\mathcal{D}}{=} Y$  does not necessarily imply that  $XZ \stackrel{\mathcal{D}}{=} YZ$ .

Define a random variable  $X$  to have a *degenerate distribution* if  $X$  is a.s. constant.

6. Show that if  $X_n \xrightarrow{\mathcal{D}} X$  and  $X$  has a degenerate distribution, then  $X_n \xrightarrow{P} X$ .

## 2. THE COMPACTNESS OF DISTRIBUTION FUNCTIONS

One of the most frequently used tools in  $\mathcal{D}$ -convergence is a certain compactness property of distribution functions. They themselves are not compact, but we can look at a slightly larger set of functions.

**Definition 8.5.** Let  $\mathcal{M}$  be the class of all functions  $G(x)$  satisfying (8.4 i and ii), with the addition of

$$(iii) \quad 0 \leq G(-\infty), \quad G(+\infty) \leq 1.$$

As before  $G, G_n \in \mathcal{M}$ ,  $G_n \xrightarrow{\mathcal{D}} G$  if  $\lim_n G_n(x) = G(x)$  at all points of  $C(G)$ .

**Theorem 8.6 (Helly-Bray).**  $\mathcal{M}$  is sequentially compact under  $\xrightarrow{\mathcal{D}}$ .

*Proof.* Let  $G_n \in \mathcal{M}$ , take  $T = \{x_k\}$ ,  $k = 1, 2, \dots$  dense in  $R^{(1)}$ . We apply Cantor's diagonalization method. That is, let  $I_1 = \{n_1, n_2, \dots\}$  be an ordered subset of the integers such that  $G_n(x_1)$  converges as  $n \rightarrow \infty$  through  $I_1$ . Let  $I_2 \subset I_1$  be such that  $G_n(x_2)$  converges as  $n \rightarrow \infty$  through  $I_2$ . Continue this way getting decreasing ordered subsets  $I_1, I_2, \dots$  of the integers. Let  $n_m$  be the  $m$ th member of  $I_m$ . For  $m \geq k$ ,  $n_m \in I_k$ , so for every  $x_k \in T$ ,  $G_{n_m}(x_k)$  converges. Define  $G_0(x)$  on  $T$  by  $G_0(x_k) = \lim_m G_{n_m}(x_k)$ . Define  $G(x)$  on  $R^{(1)}$  by

$$G(x) = \lim_{x_k \uparrow x} G_0(x_k).$$

It is easy to check that  $G \in \mathcal{M}$ . Let  $x \in C(G)$ . Then

$$\lim_{x_k \uparrow x} G_0(x_k) = G(x),$$

by definition, but also check that

$$\lim_{x_k \downarrow x} G_0(x_k) = G(x).$$

Take  $x'_k < x < x''_k$ ,  $x'_k, x''_k \in T$ . Then

$$G_{n_m}(x'_k) \leq G_{n_m}(x) \leq G_{n_m}(x''_k),$$

implying

$$G_0(x'_k) \leq \varliminf G_{n_m}(x) \leq \varlimsup G_{n_m}(x) \leq G_0(x''_k).$$

Letting  $x'_k \uparrow x$ ,  $x''_k \downarrow x$  gives the result that  $G_{n_m}$  converges to  $G$  at every  $x \in C(G)$ .

A useful way of looking at the Helly-Bray theorem is

**Corollary 8.7.** *Let  $G_n \in \mathcal{M}$ . If there is a  $G \in \mathcal{M}$  such that for every  $\mathcal{D}$ -convergent subsequence  $G_{n_m}$ ,  $G_{n_m} \xrightarrow{\mathcal{D}} G$ , then the full sequence  $G_n \xrightarrow{\mathcal{D}} G$ .*

*Proof.* If  $G_n \xrightarrow{\mathcal{D}} G$ , there exists an  $x_0 \in C(G)$  such that  $G_n(x_0) \rightarrow G(x_0)$ . But every subsequence of the  $G_n$  contains a convergent subsequence  $G_{n_m}$  and  $G_{n_m}(x_0) \rightarrow G(x_0)$ .

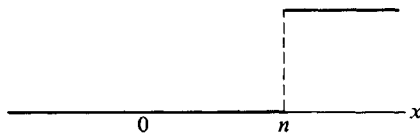


Fig. 8.1  $F_n(x)$ .

Unfortunately, the class of distribution functions itself is not compact. For instance, take  $X_n \equiv n$  (see Fig. 8.1). Obviously  $\lim_n F_n(x) = 0$  identically. The difficulty here is that mass floats out to infinity, disappearing in the limit. We want to use the Helly-Bray theorem to get some compactness properties for distribution functions. But to do this we are going to have to impose additional restrictions to keep the mass from moving out to infinity. We take some liberties with the notation.

**Definition 8.8.**  $F(B)$ ,  $B \in \mathfrak{B}_1$ , will denote the extension of  $F(x)$  to a probability measure on  $\mathfrak{B}_1$ , that is, if  $F(x) = P(X < x)$ ,  $F(B) = P(X \in B)$ .

**Definition 8.9.** Let  $\mathcal{N}$  denote the set of all distribution functions. A subset  $\mathcal{L} \subset \mathcal{N}$  will be said to be mass-preserving if for any  $\epsilon > 0$ , there is a finite interval  $I$  such that  $F(I^c) < \epsilon$ , all  $F \in \mathcal{L}$ .

**Proposition 8.10.** Let  $\mathcal{L} \subset \mathcal{N}$ . Then  $\mathcal{L}$  is conditionally compact in  $\mathcal{N}$  if and only if  $\mathcal{L}$  is mass-preserving (that is,  $F_n \in \mathcal{L} \Rightarrow \exists F_{n_m}$  such that  $F_{n_m} \xrightarrow{\mathcal{D}} F \in \mathcal{N}$ ).

*Proof.* Assume  $\mathcal{L}$  mass-preserving,  $F_n \in \mathcal{L}$ . There exists  $G \in \mathcal{M}$  such that  $F_{n_m} \xrightarrow{\mathcal{D}} G$ . For any  $\epsilon$ , take  $a, b$  such that  $F_n([a, b]) \geq 1 - \epsilon$ . Take  $a' < a$ ,  $b' > b$  so that  $a', b' \in C(G)$ . Then

$$F_{n_m}([a', b']) = F_{n_m}(b') - F_{n_m}(a') \rightarrow G(b') - G(a'),$$

with the conclusion that  $G(b') - G(a') \geq 1 - \epsilon$ , or  $G(+\infty) = 1$ ,  $G(-\infty) = 0$ , hence  $G \in \mathcal{N}$ . On the other hand, let  $\mathcal{L}$  be conditionally compact in  $\mathcal{N}$ . If  $\mathcal{L}$  is not mass-preserving, then there is an  $\epsilon > 0$  such that for every finite  $I$ ,

$$\inf_{F \in \mathcal{L}} F(I) < 1 - \epsilon.$$

Take  $F_n \in \mathcal{L}$  such that for every  $n$ ,  $F_n([-n, +n]) < 1 - \epsilon$ . Now take a subsequence  $F_{n_m} \xrightarrow{\mathcal{D}} F \in \mathcal{N}$ . Let  $a, b \in C(F)$ ; then  $F_{n_m}([a, b]) \rightarrow F([a, b])$ , but for  $n_m$  sufficiently large  $[a, b] \subset [-n_m, +n_m]$ . Thus  $F([a, b]) \leq 1 - \epsilon$  for any  $a, b \in C(F)$  which implies  $F \notin \mathcal{N}$ .

One obvious corollary of 8.10 is

**Corollary 8.11.** If  $F_n \xrightarrow{\mathcal{D}} F$ ,  $F \in \mathcal{N}$ , then  $\{F_n\}$  is mass-preserving.

## Problems

7. For  $-\infty < a < b < +\infty$ , consider the class of all distribution functions such that  $F(a) = 0$ ,  $F(b) = 1$ . Show that this class is sequentially compact.

8. Let  $F_n \xrightarrow{\mathcal{D}} F$ , and  $F_n, F$  be distribution functions. Show that for  $B$ , any Borel set, it is not necessarily true that  $F_n(B) = 1$ , for all  $n \Rightarrow F(B) = 1$ . Show that if  $B$  is closed, however, then  $F_n(B) = 1$ , for all  $n \Rightarrow F(B) = 1$ .

9. Let  $g(x)$  be  $\mathfrak{B}_1$ -measurable, such that  $|g(x)| \rightarrow \infty$  as  $x \rightarrow \pm\infty$ . If  $\mathcal{L} \subset \mathcal{N}$  is such that  $\sup_{F \in \mathcal{L}} \int |g| dF < \infty$ , then  $\mathcal{L}$  is mass-preserving.

10. Show that if there is an  $r > 0$  such that  $\overline{\lim} E|X_n|^r < \infty$ , then  $\{F_n\}$  is mass-preserving.

11. The support of  $F$  is the smallest closed set  $C$  such that  $F(C) = 1$ . Show that such a minimal closed set exists. A point of increase of  $F$  is

defined as a point  $x$  such that for every neighborhood  $N$  of  $x$ ,  $F(N) > 0$ . Show that the set of all points of increase is exactly the support of  $F$ .

**12.** Define a Markov chain with stationary transition probabilities  $p(\cdot | x)$  to be *stable* if for any sequence of initial distributions  $\pi_n$   $\mathcal{D}$ -converging to an initial distribution  $\pi$ , the probabilities  $\int p(\cdot | x)\pi_n(dx)$   $\mathcal{D}$ -converge to the probability  $\int p(\cdot | x)\pi(dx)$ .

If the state space of a stable Markov chain is a compact interval, show that there is at least one invariant initial distribution. [Use Problem 7 applied to the probabilities  $1/n \sum_1^n p^{(k)}(\cdot | x)$  for  $x$  fixed.]

### 3. INTEGRALS AND $\mathcal{D}$ -CONVERGENCE

Suppose  $F_n \xrightarrow{\mathcal{D}} F$ ,  $F_n, F \in \mathcal{N}$ , does it then follow that for any reasonable measurable function  $f(x)$ , that  $\int f(x) dF_n \rightarrow \int f(x) dF$ ? The answer is No! For example, let

$$X_n \equiv 1/n, X \equiv 0, \text{ so } F_n \xrightarrow{\mathcal{D}} F.$$

Now take  $f(x) = 0$ ,  $x \leq 0$ , and  $f(x) = 1$ ,  $x > 0$ . Then  $\int f dF_n = 1$ , but  $\int f dF = 0$ . But it is easy to see that it works for  $f$  bounded and continuous. Actually, a little more can be said.

**Proposition 8.12.** Let  $F_n, F \in \mathcal{N}$  and  $F_n \xrightarrow{\mathcal{D}} F$ . If  $f(x)$  is bounded on  $R^{(1)}$ , measurable  $\mathcal{B}_1$  and the discontinuity points of  $f$  are a set  $S$  with  $F(S) = 0$ , then

$$\int f dF_n \rightarrow \int f dF.$$

**Remark.** The set of discontinuity points of a  $\mathcal{B}_1$ -measurable function is in  $\mathcal{B}_1$ , so  $F(S)$  is well-defined. (See Hobson [72, p. 313].)

*Proof.* Take  $a, b \in C(F)$ ,  $I_1, \dots, I_k$  a partition  $\mathcal{P}_k$  of  $I = [a, b]$ , where  $I_i = [a_i, b_i]$  and  $a_i, b_i \in C(F)$ . Define on  $I$

$$f_k^+(x) = \sum_1^k \sup_{x \in I_i} (f(x)) \chi_{I_i}(x),$$

$$f_k^-(x) = \sum_1^k \inf_{x \in I_i} (f(x)) \chi_{I_i}(x).$$

Then

$$\int_I f_k^- dF_n \leq \int_I f dF_n \leq \int_I f_k^+ dF_n.$$

Clearly, the right- and left-hand sides above converge, and

$$\int_I f_k^- dF \leq \varliminf \int_I f dF_n \leq \varlimsup \int_I f dF_n \leq \int_I f_k^+ dF.$$



Let  $\|\mathcal{F}_k\| \rightarrow 0$ . At every point  $x$  which is a continuity point of  $f(x)$ ,

$$f_k^-(x) \uparrow f(x), \quad f_k^+(x) \downarrow f(x).$$

By the Lebesgue bounded convergence theorem,

$$\lim_k \int_I f_k^- dF = \lim_k \int_I f_k^+ dF = \int_I f dF.$$

Let  $M = \sup |f(x)|$ , then since  $\{F_n\}$  is mass-preserving, for any  $\epsilon > 0$  we can take  $I$  such that  $F_n(I^c) \leq \epsilon/2M$  and  $F(I^c) \leq \epsilon/2M$ . Now

$$\overline{\lim} \left| \int f dF_n - \int f dF \right| \leq \overline{\lim} \left| \int_I f dF_n - \int_I f dF \right| + \epsilon = \epsilon.$$

**Corollary 8.13.** *In the above proposition, eliminate the condition that  $f$  be bounded on  $R^{(1)}$ , then*

$$\int |f| dF \leq \underline{\lim} \int |f| dF_n.$$

*Proof.* Define

$$g_\alpha(x) = \begin{cases} |f(x)|, & \text{if } |f(x)| \leq \alpha, \\ \alpha, & \text{otherwise.} \end{cases}$$

Every continuity point of  $f$  is a continuity point of  $g_\alpha$ . Apply Proposition 8.12 to  $g_\alpha$  to conclude

$$\int g_\alpha dF = \lim \int g_\alpha dF_n \leq \underline{\lim} \int |f| dF_n.$$

Let  $\alpha \uparrow \infty$ , then  $g_\alpha \uparrow |f|$ . By the monotone convergence theorem  $\int g_\alpha dF \rightarrow \int |f| dF$ .

### Problems

**13.** Let  $F_n, F \in \mathcal{N}$ ,  $F_n \xrightarrow{\mathcal{D}} F$ . For any set  $E \subset R^{(1)}$ , define the boundary of  $E$  as  $bd(E) = \bar{E} \cap \bar{E}^c$  ( $\bar{E}$  = closure of  $E$ ). Prove that for any  $B \in \mathcal{B}_1$  such that  $F(bd(B)) = 0$ ,  $F_n(B) \rightarrow F(B)$ .

**14.** Let  $F_n \xrightarrow{\mathcal{D}} F$ , and  $h(x), g(x)$  be continuous functions such that

$$\begin{aligned} |g(x)| &\rightarrow +\infty \quad \text{as } x \rightarrow \pm\infty, \\ \left| \frac{h(x)}{g(x)} \right| &\rightarrow 0 \quad \text{as } x \rightarrow \pm\infty. \end{aligned}$$

Show that  $\overline{\lim} \int |g(x)| dF_n < \infty$  implies

$$\int h(x) dF_n \rightarrow \int h(x) dF.$$

#### 4. CLASSES OF FUNCTIONS THAT SEPARATE

**Definition 8.14.** A set  $\mathcal{E}$  of bounded continuous functions on  $R^{(1)}$  will be called  $\mathcal{N}$ -separating if for any  $F, G \in \mathcal{N}$ ,

$$\int f dF = \int f dG, \quad \text{all } f \in \mathcal{E}$$

implies  $F = G$ .

We make this a bit more general (also ultimately, more convenient) by allowing the functions of  $\mathcal{E}$  to be complex-valued. That is, we consider functions  $f(x)$  of the form  $f(x) = f_1(x) + if_2(x)$ ,  $f_1, f_2$  real-valued, continuous, and bounded. Now, of course,  $|f(x)|$  has the meaning of the absolute value of a complex number. As usual, then,  $\int f dF = \int f_1 dF + i \int f_2 dF$ .

The nice thing about such a class  $\mathcal{E}$  of functions is that we can check whether  $F_n \xrightarrow{\mathcal{D}}$  by looking at the integrals of these functions. More specifically:

**Proposition 8.15.** Let  $\mathcal{E}$  be  $\mathcal{N}$ -separating, and  $\{F_n\}$  mass-preserving. Then there exists an  $F \in \mathcal{N}$  such that  $F_n \xrightarrow{\mathcal{D}} F$  if and only if

$$\lim_n \int f dF_n \text{ exists, all } f \in \mathcal{E}.$$

If this holds, then  $\lim_n \int f dF_n = \int f dF$ , all  $f \in \mathcal{E}$ .

*Proof.* One way is clear. If  $F_n \xrightarrow{\mathcal{D}} F$ , then  $\lim_n \int f dF_n = \int f dF$  by 8.12.

To go the other way, take any  $\mathcal{D}$ -convergent subsequence  $F_{n_k}$  of  $F_n$ . By mass-preservation  $F_{n_k} \xrightarrow{\mathcal{D}} F \in \mathcal{N}$ . Take any other convergent subsequence  $F_{n'_k} \xrightarrow{\mathcal{D}} G$ . Then for  $f \in \mathcal{E}$ , by 8.12,

$$\lim_k \int f dF_{n_k} = \int f dF, \quad \lim_k \int f dF_{n'_k} = \int f dG,$$

so  $\int f dF = \int f dG$ , all  $f \in \mathcal{E} \Rightarrow F = G$ . All  $\mathcal{D}$ -convergent subsequences of  $F_n$  have the same limit  $F$ , implying  $F_n \xrightarrow{\mathcal{D}} F$ .

**Corollary 8.16.** Let  $\mathcal{E}$  be  $\mathcal{N}$ -separating and  $\{F_n\}$  mass-preserving. If  $F \in \mathcal{N}$  is such that  $\int f dF_n \rightarrow \int f dF$ , all  $f \in \mathcal{E}$ , then  $F_n \xrightarrow{\mathcal{D}} F$ .

The relevance of looking at integrals of functions to  $\mathcal{D}$ -convergence can be clarified by the simple observation that  $F_n \xrightarrow{\mathcal{D}} F$  is equivalent to  $\int f dF_n \rightarrow \int f dF$  for all functions  $f$  of the form

$$f(x) = \begin{cases} 1, & x < x_0, \\ 0, & x \geq x_0, \end{cases}$$

for any  $x_0 \in C(F)$ .

What classes of functions are  $\mathcal{N}$ -separating? Take  $\mathcal{E}_0$  to be all functions  $f$  of the form below (see Fig. 8.2) with any  $a, b$  finite and any  $\epsilon > 0$ .

**Proposition 8.17.**  $\mathcal{E}_0$  is  $\mathcal{N}$ -separating.

*Proof.* For any  $F, G \in \mathcal{N}$ , take  $a, b \in C(F) \cap C(G)$ . Assume that for any  $f$  as described,

$$\int f dF = \int f dG.$$

Then

$$F([a, b]) \leq \int f dF = \int f dG \leq G([a - \epsilon, b + \epsilon]),$$

and conversely,

$$G([a, b]) \leq F([a - \epsilon, b + \epsilon]).$$

Let  $\epsilon \downarrow 0$ , to get  $F([a, b]) = G([a, b])$ . The foregoing being true for all  $a, b \in C(F) \cap C(G)$  implies  $F = G$ .

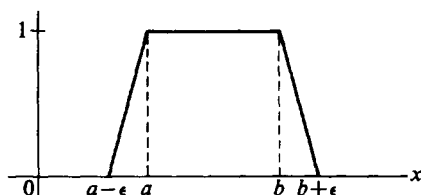


Figure 8.2

However  $\mathcal{E}_0$  is an awkward set of functions to work with. What is really more important is

**Proposition 8.18.** Let  $\mathcal{E}$  be a class of continuous bounded functions on  $R^{(1)}$  with the property that for any  $f_0 \in \mathcal{E}_0$ , there exist  $f_n \in \mathcal{E}$  such that  $\sup |f_n(x)| \leq M$ , all  $n$ , and  $\lim_n f_n(x) = f_0(x)$  for every  $x \in R^{(1)}$ . Then  $\mathcal{E}$  is  $\mathcal{N}$ -separating.

*Proof.* Let  $\int f dF = \int f dG$ , all  $f \in \mathcal{E}$ . For any  $f_0 \in \mathcal{E}_0$ , take  $f_n \in \mathcal{E}$  converging to  $f_0$  as in the statement 8.18 above. By the Lebesgue bounded convergence theorem

$$\lim_n \int f_n dF = \int f_0 dF, \quad \lim_n \int f_n dG = \int f_0 dG,$$

so  $\int f_0 dF = \int f_0 dG$ , all  $f_0 \in \mathcal{E}_0$  implying  $F = G$ .

## 5. TRANSLATION INTO RANDOM-VARIABLE TERMS

The foregoing is all translatable into random-variable terms. For example:

- i) If  $X_n$  are random variables, their distribution functions are mass-preserving iff for any  $\epsilon > 0$ , there is a finite interval  $I$  such that

$$P(X_n \in I^c) < \epsilon, \quad \text{all } n.$$

- ii) If  $|g(x)| \rightarrow \infty$  as  $x \rightarrow \pm\infty$ , then the distribution functions of  $X_n$  are mass-preserving if  $\sup E |g(X_n)| < \infty$  (Problem 9).
- iii) If  $X_n$  have mass-preserving distribution functions and  $\mathcal{E}$  is an  $\mathcal{N}$ -separating set of functions, then there exists a random variable  $X$  such that  $X_n \xrightarrow{\mathcal{D}} X$  if and only if  $\lim_n Ef(X_n)$  exists, all  $f \in \mathcal{E}$ .

We will switch freely between discussion in terms of distribution functions and in terms of random variables, depending on which set of terms is more illuminating.

**Proposition 8.19.** Let  $X_n \xrightarrow{\mathcal{D}} X$ , and let  $\varphi(x)$  be measurable  $\mathcal{B}_1$ , with its set  $S$  of discontinuities such that  $P(X \in S) = 0$ . Then

$$\varphi(X_n) \xrightarrow{\mathcal{D}} \varphi(X).$$

*Proof.* Let  $Z_n = \varphi(X_n)$ ,  $Z = \varphi(X)$ . If  $Ef(Z_n) \rightarrow Ef(Z)$ , for all  $f \in \mathcal{E}_0$ , then  $Z_n \xrightarrow{\mathcal{D}} Z$ . Let  $g(x) = f(\varphi(x))$ . This function  $g$  is bounded, measurable  $\mathcal{B}_1$ , and continuous wherever  $\varphi$  is continuous. By 8.12,  $Eg(X_n) \rightarrow Eg(X)$ .

We can't do any better with a.s. convergence. This is illustrated by the following problem.

**Problem 15.** If  $\varphi(x)$  is as in 8.19, and  $X_n \xrightarrow{\text{a.s.}} X$ , then show  $\varphi(X_n) \xrightarrow{\text{a.s.}} \varphi(X)$ . Give an example to show that in general this is not true if  $\varphi(x)$  is only assumed measurable.

## 6. AN APPLICATION OF THE FOREGOING

With only this scanty background we are already in a position to prove a more general version of the central limit theorem. To do this we work with the class of functions defined by

$\mathcal{E}_1$  consists of all continuous bounded  $f$  on  $R^{(1)}$  such that  $f''(x)$  exists for all  $x$ ,  $\sup_x |f''(x)| < \infty$ , and  $f''(x)$  is uniformly continuous on  $R^{(1)}$ .

It is fairly obvious that  $\mathcal{E}_1$  satisfies the requirements of 8.18 and hence is  $\mathcal{N}$ -separating. We use  $\mathcal{E}_1$  to establish a simple example of what has become known as the "invariance principle."

**Theorem 8.20.** If there is one sequence  $X_1^*, X_2^*, \dots$  of independent, identically distributed random variables,  $EX_1^* = 0$ ,  $E(X_1^*)^2 = \sigma^{*2} < \infty$ , such that

$$\frac{X_1^* + \dots + X_n^*}{\sigma^* \sqrt{n}} \xrightarrow{\mathcal{D}} X,$$

then for all sequences  $X_1, X_2, \dots$  of independent, identically distributed random

variables such that  $EX_1 = 0$ ,  $EX_1^2 = \sigma^2 < \infty$ ,

$$\frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} X.$$

*Proof.* Let  $f \in \mathcal{E}_1$ , and define

$$\delta(h) = \sup_{|x-y| \leq h} |f''(x) - f''(y)|.$$

By definition  $\lim_{h \downarrow 0} \delta(h) = 0$  as  $h \downarrow 0$ . We may as well assume  $\sigma^* = \sigma = 1$ , otherwise we use  $X_k^*/\sigma^*$ ,  $X_k/\sigma$ . Let

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}, \quad Z_n^* = \frac{X_1^* + \cdots + X_n^*}{\sqrt{n}}.$$

Since  $EZ_n^2 = 1$ ,  $E(Z_n^*)^2 = 1$ , both sequences are mass-preserving. By 8.15

$$Ef(Z_n^*) \rightarrow Ef(X), \quad \text{all } f \in \mathcal{E}_1,$$

and by 8.16 it suffices to show that

$$Ef(Z_n) \rightarrow Ef(X), \quad \text{all } f \in \mathcal{E}_1.$$

Since only the distributions are relevant here, we can assume that  $X^*$ ,  $X$  are defined on a common sample space and are independent of each other. Now write

$$\begin{aligned} f(Z_n) - f(Z_n^*) &= f\left(\frac{X_1 + \cdots + X_{n-1} + X_n}{\sqrt{n}}\right) - f\left(\frac{X_1 + \cdots + X_{n-1} + X_n^*}{\sqrt{n}}\right) \\ &+ f\left(\frac{X_1 + \cdots + X_{n-1} + X_n^*}{\sqrt{n}}\right) - f\left(\frac{X_1 + \cdots + X_{n-2} + X_{n-1}^* + X_n^*}{\sqrt{n}}\right) \\ &+ \cdots + f\left(\frac{X_1 + X_2^* + \cdots + X_n^*}{\sqrt{n}}\right) - f\left(\frac{X_1^* + \cdots + X_n^*}{\sqrt{n}}\right). \end{aligned}$$

Define random variables  $V_k, U_k$  by

$$\begin{aligned} V_k &= f\left(\frac{X_1 + \cdots + X_k + X_{k+1}^* + \cdots + X_n^*}{\sqrt{n}}\right) \\ &\quad - f\left(\frac{X_1 + \cdots + X_{k-1} + X_k^* + \cdots + X_n^*}{\sqrt{n}}\right), \\ U_k &= \frac{X_1 + \cdots + X_{k-1} + X_{k+1}^* + \cdots + X_n^*}{\sqrt{n}}. \end{aligned}$$

Use Taylor's expansion around  $U_k$  to get

$$\begin{aligned}
 (8.21) \quad V_k &= f\left(U_k + \frac{X_k}{\sqrt{n}}\right) - f\left(U_k + \frac{X_k^*}{\sqrt{n}}\right) \\
 &= \frac{X_k - X_k^*}{\sqrt{n}} f'(U_k) + \frac{X_k^2}{2n} f''\left(U_k + \frac{\theta X_k}{\sqrt{n}}\right) \\
 &\quad - \frac{X_k^{*2}}{2n} f''\left(U_k + \frac{\theta^* X_k^*}{\sqrt{n}}\right),
 \end{aligned}$$

where  $\theta, \theta^*$  are random variables such that  $0 \leq \theta, \theta^* \leq 1$ . Both  $X_k, X_k^*$  are independent of  $U_k$ , so

$$E[(X_k - X_k^*)f'(U_k)] = 0, \quad (|f'(x)| \leq c_0 + c_1|x| \Rightarrow E|f'(U_k)| < \infty).$$

$$E\left(X_k^2 f''\left(U_k + \frac{\theta X_k}{\sqrt{n}}\right)\right) = (EX_k^2)(Ef''(U_k)) + \alpha E\left(X_k^2 \delta\left(\frac{|X_k|}{\sqrt{n}}\right)\right), \quad |\alpha| \leq 1.$$

Let  $h_n(x) = x^2 \delta(|x|/\sqrt{n})$ . Take the expectation of (8.21) and use  $EX_k^2 = E(X_k^*)^2$  to get

$$|EV_k| \leq (1/2n)[Eh_n(X_k) + Eh_n(X_k^*)] = (1/2n)[Eh_n(X_1) + Eh_n(X_1^*)],$$

this latter by the identical distribution. Note that

$$f(Z_n) - f(Z_n^*) = V_n + \cdots + V_1,$$

so

$$|Ef(Z_n) - Ef(Z_n^*)| \leq \frac{1}{2}Eh_n(X_1) + \frac{1}{2}Eh_n(X_1^*)$$

Let  $M = \sup_x |f''(x)|$ ; then  $\delta(h) \leq 2M$ , all  $h$ , so  $h_n(X_1) \leq 2MX_1^2$ . But  $h_n(X_1) \rightarrow 0$  a.s. Since  $X_1^2$  is integrable, the bounded convergence theorem yields  $Eh_n(X_1) \rightarrow 0$ . Similarly for  $Eh_n(X_1^*)$ . Thus, it has been established that

$$|Ef(Z_n) - Ef(Z_n^*)| \rightarrow 0$$

implying

$$Ef(Z_n) \rightarrow Ef(X). \quad \text{Q.E.D.}$$

This proof is anachronistic in the sense that there are much simpler methods of proving the central limit theorem if one knows some more probability theory. But it is an interesting proof. We know that if we take  $X_1^*, X_2^*, \dots$  to be fair coin-tossing variables, that

$$\frac{X_1^* + \cdots + X_n^*}{\sqrt{n}} \rightarrow \mathcal{N}(0, 1),$$

where the notation  $\mathcal{N}(0,1)$  is clarified by

**Definition 8.22.** The normal distribution with mean  $\mu$  and variance  $\sigma^2$ , denoted  $\mathcal{N}(\mu, \sigma)$ , is the distribution of a random variable  $\sigma X + \mu$ , where

$$P(X < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

So we have proved

**Corollary 8.23.** Let  $X_1, X_2, \dots$  be independent, identically distributed random variables,  $EX_1 = 0$ ,  $EX_1^2 = \sigma^2 < \infty$ . Then

$$\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

## 7. CHARACTERISTIC FUNCTIONS AND THE CONTINUITY THEOREM

The class of functions of the form  $\{e^{iux}\}$ ,  $u \in R^{(1)}$ , is particularly important and useful in studying convergence in distribution. To begin with

**Theorem 8.24.** The set of all complex exponentials  $\{e^{iux}\}$ ,  $u \in R^{(1)}$ , is  $\mathcal{N}$ -separating.

*Proof.* Let  $\int e^{iux} dF = \int e^{iux} dG$ , all  $u$ . Then for  $\alpha_k$ ,  $k = 1, \dots, m$  any complex numbers, and  $u_1, \dots, u_m$  real,

$$(8.25) \quad \int \left( \sum_k \alpha_k e^{iu_k x} \right) dF = \int \left( \sum_k \alpha_k e^{iu_k x} \right) dG.$$

Let  $f_0$  be in  $\mathcal{E}_0$ , let  $\epsilon_n \downarrow 0$ ,  $\epsilon_n \leq 1$ , and consider the interval  $[-n, +n]$ . Any continuous function on  $[-n, +n]$  equal at endpoints can be uniformly approximated by a trigonometric polynomial; that is, there exists a finite sum

$$f_n = \sum_k a_k e^{i\pi k x/n}$$

such that  $|f_0(x) - f_n(x)| \leq \epsilon_n$ ,  $x \in [-n, +n]$ . Since  $f_n$  is periodic, and  $\epsilon_n \leq 1$ , then for all  $n$ ,  $\sup_x |f_n(x)| \leq 2$ . By (8.25) above  $\int f_n dF = \int f_n dG$ . This gives  $\int f_0 dF = \int f_0 dG$  or  $F = G$ .

**Definition 8.26.** Given a distribution function  $F(x)$ , its characteristic function  $f(u)$  is a complex-valued function defined on  $R^{(1)}$  by

$$f(u) = \int e^{iux} F(dx).$$

If  $F$  is the distribution function of the random variable  $X$ , then equivalently,

$$f(u) = E e^{iuX}.$$

Note quickly that

**Proposition 8.27.** *Any characteristic function  $f(u)$  has the properties*

- i)  $f(0) = 1$ ,
- ii)  $|f(u)| \leq 1$ ,
- iii)  $f(u)$  is uniformly continuous on  $R^{(1)}$ ,
- iv)  $f(-u) = \overline{f(u)}$ .

*Proof*

- i) Obvious;

$$\text{ii) } \left| \int e^{iux} dF \right| \leq \int |e^{iux}| dF = 1;$$

$$\text{iii) } |f(u+h) - f(u)| = \left| \int e^{iux}(e^{ixh} - 1) dF \right| \leq \int |e^{ixh} - 1| dF = \delta(h),$$

by the bounded convergence theorem  $\delta(h) \rightarrow 0$  as  $h \rightarrow 0$ ;

- iv)  $f(-u) = \overline{f(u)}$  is obvious.

Theorem 8.24 may be stated as: No two distinct distribution functions have the same characteristic function. However, examples are known (see Loève [108, p. 218]) of distribution functions  $F_1 \neq F_2$  such that  $f_1(u) = f_2(u)$  for all  $u$  in the interval  $[-1, +1]$ . Consequently, the set of functions  $\{e^{iux}\}$ ,  $-1 \leq u \leq +1$ , is not  $\mathcal{N}$ -separating.

The condition that  $F_n \xrightarrow{\mathcal{D}} F$  can be elegantly stated in terms of the associated characteristic functions.

**Theorem 8.28 (The continuity theorem).** *If  $F_n$  are distribution functions with characteristic functions  $f_n(u)$  such that*

- a)  $\lim_n f_n(u)$  exists for every  $u$ , and
- b)  $\lim_n f_n(u) = h(u)$  is continuous at  $u = 0$ ,

*then there is a distribution function  $F$  such that  $F_n \xrightarrow{\mathcal{D}} F$  and  $h(u)$  is the characteristic function of  $F$ .*

*Proof.* Since  $\{e^{iux}\}$  is  $\mathcal{N}$ -separating and  $\lim_n \int e^{iux} dF_n$  exists for every member of  $\{e^{iux}\}$ , by 8.15, all we need to do is show that  $\{F_n\}$  is mass-preserving. To do this, we need

**Proposition 8.29.** *There exists a constant  $\alpha$ ,  $0 < \alpha < \infty$ , such that for any distribution function  $F$  with characteristic function  $f$ , and any  $u > 0$ ,*

$$F\left(\left[-\frac{1}{u}, +\frac{1}{u}\right]^c\right) \leq \frac{\alpha}{u} \int_0^u (1 - Rf(v)) dv.$$



*Proof.*  $Rlf(u) = \int \cos ux F(dx)$ , so

$$\begin{aligned} \frac{1}{u} \int_0^u [1 - Rlf(v)] dv &= \frac{1}{u} \int_0^u \int (1 - \cos vx) F(dx) dv \\ &= \int \left[ \frac{1}{u} \int_0^u (1 - \cos vx) dv \right] F(dx) \\ &= \int \left( 1 - \frac{\sin ux}{ux} \right) F(dx) \\ &\geq \int_{\{|xu| \geq 1\}} \left( 1 - \frac{\sin ux}{ux} \right) F(dx) \\ &\geq \inf_{|t| \geq 1} \left( 1 - \frac{\sin t}{t} \right) F \left( \left[ -\frac{1}{u}, +\frac{1}{u} \right]^c \right). \end{aligned}$$

Letting

$$\alpha = \left[ \inf_{|t| \geq 1} \left( 1 - \frac{\sin t}{t} \right) \right]^{-1}$$

does it.

Now back to the main theorem. By the above inequality,

$$\overline{\lim} F_n \left( \left[ -\frac{1}{u}, +\frac{1}{u} \right]^c \right) \leq \frac{\alpha}{u} \overline{\lim} \int_0^u [1 - Rlf_n(v)] dv.$$

The bounded convergence theorem gives

$$\overline{\lim} F_n \left( \left[ -\frac{1}{u}, +\frac{1}{u} \right]^c \right) \leq \frac{\alpha}{u} \int_0^u (1 - Rlh(v)) dv.$$

Now  $f_n(0) = 1 \Rightarrow h(0) = 1$ . By continuity of  $h$  at zero,  $\lim_{v \rightarrow 0} Rlh(v) = 1$ . Therefore,

$$\lim_{u \rightarrow 0} \overline{\lim} F_n \left( \left[ -\frac{1}{u}, +\frac{1}{u} \right]^c \right) = 0.$$

By this, for any  $\epsilon > 0$  we may take  $a$  so that  $\overline{\lim} F_n([-a, +a]^c) < \epsilon/2$ . So there is an  $n_0$  such that for  $n > n_0$ ,  $F_n([-a, +a]^c) < \epsilon$ . Take  $b > a$  such that  $F_k([-b, +b]^c) < \epsilon$  for  $k = 1, 2, \dots, n_0$ . From these together  $\sup_n F_n([-b, +b]^c) < \epsilon$ . Q.E.D.

**Corollary 8.30.** Let  $F_n$  be distribution functions,  $f_n$  their characteristic functions. If there is a distribution function  $F$  with characteristic function  $f$  such that  $\lim_n f_n(u) = f(u)$  for every  $u$ , then  $F_n \xrightarrow{D} F$ .

*Proof.* Obvious from 8.28.

Clearly, if  $F_n \xrightarrow{\mathcal{D}} F$ , then the characteristic functions  $f_n(u)$  converge at every point  $u$  to  $f(u)$ . We strengthen this to

**Proposition 8.31.** *If  $F_n \xrightarrow{\mathcal{D}} F$ , then the corresponding characteristic functions  $f_n(u)$  converge uniformly to  $f(u)$  on every finite interval  $I$ . (Denote this kind of convergence by  $\xrightarrow{uc}$ ).*

*Proof.* This result follows from the fact that the  $f_n, f$  form an equicontinuous family. That is, if we fix a finite interval  $I$ , then for any  $n, u$ , and  $h$ ,

$$\begin{aligned} |f_n(u+h) - f_n(u)| &\leq \left| \int_I (e^{i(u+h)x} - e^{iux}) dF_n(x) \right| + 2F_n(I^c) \\ &\leq \sup_{x \in I} |e^{ihx} - 1| + 2F_n(I^c). \end{aligned}$$

Thus, since the  $\{F_n\}$  are mass-preserving,

$$\sup_n |f_n(u+h) - f_n(u)| \leq \delta(h),$$

where  $\delta(h) \downarrow 0$  as  $h \downarrow 0$ . Now the usual argument works: Divide  $I$  up into points  $u_1, u_2, \dots, u_m$  such that  $|u_{k+1} - u_k| \leq h$ . For  $u \in I$ ,

$$|f_n(u) - f(u)| \leq |f_n(u) - f_n(u_k)| + |f(u) - f(u_k)| + |f_n(u_k) - f(u_k)|,$$

where  $u_k$  is the point of the partition nearest  $u$ . Therefore

$$\limsup_n \sup_{u \in I} |f_n(u) - f(u)| \leq 2\delta(h),$$

because  $f(u)$  also satisfies  $|f(u+h) - f(u)| \leq \delta(h)$ . Taking  $h \rightarrow 0$  now completes the proof.

The continuity theorem gives us a strong basic tool. Now we start reaping limit theorems from it by using some additional technical details.

### Problems

**16.** A random variable  $X$  has a symmetric distribution if  $P(X \in B) = P(X \in -B)$ , where  $-B = \{-x; x \in B\}$ . Prove that the characteristic function of  $X$  is real for all  $u$  iff  $X$  has a symmetric distribution.

**17.** A natural question is, what continuous complex-valued functions  $f(u)$  on  $R^{(1)}$  are characteristic functions? Say that such a function is nonnegative definite if for any complex numbers  $\lambda_1, \dots, \lambda_n$ , and points  $u_1, \dots, u_n \in R^{(1)}$ ,

$$\sum_{i,j=1}^n f(u_i - u_j) \lambda_i \bar{\lambda}_j \geq 0.$$

A complete answer to the question is given by the following theorem.

**Bochner's Theorem.** Let  $f(u)$  be continuous on  $R^{(1)}$ ,  $f(0) = 1$ . Then  $f$  is a characteristic function if and only if it is nonnegative definite.

Prove that if  $f$  is a characteristic function, then it is nonnegative definite. (See Loève [108, pp. 207 ff.] for a proof of the other direction.)

18. Find the characteristic function for a Poisson distribution with parameter  $\lambda$ .
19. Find the characteristic function of  $S_n$  for coin-tossing.
20. If  $Y = aX + b$ , show that

$$f_Y(u) = e^{iub} f_X(au).$$

21. A random variable  $X$  is called a displaced lattice random variable if there are numbers  $a, d$  such that

$$\sum_{n=-\infty}^{+\infty} P(X = nd + a) = 1.$$

Show that  $X$  is a displaced lattice if and only if there is a  $u \neq 0$  such that  $|f_X(u)| = 1$ . If  $u_1, u_2$  are irrational with respect to each other, and  $|f_X(u_1)| = |f_X(u_2)| = 1$ , show that  $X$  is a.s. constant, hence  $|f_X(u)| \equiv 1$ . Show that  $X$  is distributed on a lattice  $L_d$ ,  $d > 0$  iff there is a  $u \neq 0$  such that  $f_X(u) = 1$ .

## 8. THE CONVERGENCE OF TYPES THEOREM

Look at the question: Suppose that  $X_n \xrightarrow{\mathcal{D}} X$ , and  $X$  is nondegenerate. Can we find constants  $a_n, b_n$  such that  $a_n X_n + b_n \xrightarrow{\mathcal{D}} X'$  where  $X'$  has a law not connected with that of  $X$  in any reasonable way? For example, if  $X_1, X_2, \dots$  are independent and identically distributed,  $EX_1 = 0, EX_1^2 < \infty$ , can we find constants  $\lambda_n$  such that  $S_n/\lambda_n$   $\mathcal{D}$ -converges to something not  $\mathcal{N}(\mu, \sigma)$ ? And if  $S_n/\lambda_n$   $\mathcal{D}$ -converges, what can be said about the size of  $\lambda_n$  compared with  $\sqrt{n}$ , the normalizing factor we have been using? Clearly, we cannot get the result that  $X_n \xrightarrow{\mathcal{D}} X, a_n X_n + b_n \xrightarrow{\mathcal{D}} X'$  implies  $\lim a_n = a$  exists, because if  $X_n$  has a symmetric distribution, then  $X_n \xrightarrow{\mathcal{D}} X \Rightarrow (-1)^n X_n \xrightarrow{\mathcal{D}} X$ , since  $-X_n$  and  $X_n$  have the same law. But if we rule this out by requiring  $a_n > 0$ , then the kind of result we want holds.

**Theorem 8.32 (Convergence of types theorem).** Let  $X_n \xrightarrow{\mathcal{D}} X$ , and suppose there are constants  $a_n > 0, b_n$  such that  $a_n X_n + b_n \xrightarrow{\mathcal{D}} X'$ , where  $X$  and  $X'$  are nondegenerate. Then there are constants  $a, b$  such that  $\mathbb{L}(X') = \mathbb{L}(aX + b)$  and  $b_n \rightarrow b, a_n \rightarrow a$ .

*Proof.* Use characteristic functions and let  $f_n = f_{X_n}$  so that

$$f_{a_n X_n + b_n}(u) = e^{iub_n} f_n(a_n u).$$

By 8.31, if  $f', f$  are the characteristic functions of  $X', X$  respectively, then

$$e^{iub_n} f_n(a_n u) \xrightarrow{uc} f'(u), \quad f_n(u) \xrightarrow{uc} f(u).$$

Take  $n_m$  such that  $a_{n_m} \rightarrow a$ , where  $a$  may be infinite. Since

$$|f_n(a_n u)| \xrightarrow{uc} |f'(u)|, \quad |f_n(u)| \xrightarrow{uc} |f(u)|,$$

if  $a_n \rightarrow \infty$ , substitute  $v_n = u/a_n$ ,  $u \in I$ , to get

$$|f_n(u)| = |f_n(a_n v_n)| \rightarrow |f'(0)|.$$

Thus  $|f(u)| \equiv 1$ , implying  $X$  degenerate by Problem 21. Hence  $a$  is finite. Using  $uc$ -convergence  $|f_n(a_n u)| \rightarrow |f(au)|$ ; thus  $|f'(u)| = |f(au)|$ . Suppose  $a_{n_m} \rightarrow a$ ,  $a_{n'_m} \rightarrow a'$  and  $a \neq a'$ . Use  $|f(au)| = |f(a'u)|$ , assume  $a' < a$ , so  $|f(u)| = |f((a'/a)u)| = \cdots = |f((a'/a)^N u)|$  by iterating  $N$  times. Let  $N \rightarrow \infty$  to get the contradiction  $|f(u)| \equiv 1$ . Thus there is a unique  $a \geq 0$  such that  $a_n \rightarrow a$ . So  $f_n(a_n u) \rightarrow f(au)$ . Hence  $e^{iub_n}$  must converge for every  $u$  such that  $f(au) \neq 0$ , thus in some interval  $|u| < \delta$ . Obviously then,  $\lim |b_n| < \infty$ , and if  $b, b'$  are two limit-points of  $b_n$ , then  $e^{iub} = e^{iub'}$  for all  $|u| < \delta$ , which implies  $b = b'$ . Thus  $b_n \rightarrow b$ ,  $e^{iub_n} \rightarrow e^{iub}$ , and  $f'(u) = e^{iub} f(au)$ .

## 9. CHARACTERISTIC FUNCTIONS AND INDEPENDENCE

The part that is really important and makes the use of characteristic functions so natural is the multiplicative property of the complex exponentials and the way that this property fits in with the independence of random variables.

**Proposition 8.33.** Let  $X_1, X_2, \dots, X_n$  be random variables with characteristic functions  $f_1(u), \dots, f_n(u)$ . The random variables are independent iff for all  $u_1, \dots, u_n$ ,

$$E\left(\exp\left[i\sum_{k=1}^n u_k X_k\right]\right) = \prod_{k=1}^n f_k(u_k).$$

*Proof.* Suppose  $X, Y$  are independent random variables and  $f, g$  are complex-valued measurable functions,  $f = f_1 + if_2$ ,  $g = g_1 + ig_2$ , and  $f_1, f_2, g_1, g_2$  are  $\mathcal{B}_1$ -measurable. Then I assert that if  $E|f(X)| < \infty$ ,  $E|g(Y)| < \infty$ ,

$$(8.34) \quad E(f(X)g(Y)) = (Ef(X))(Eg(Y)),$$

so splitting into products does carry over to complex-valued functions. To show this, just verify

$$\begin{aligned} E[(f_1(X) + if_2(X))(g_1(Y) + ig_2(Y))] \\ = Ef_1(X)g_1(Y) + iEf_2(X)g_1(Y) + iEf_1(X)g_2(Y) - Ef_2(X)g_2(Y). \end{aligned}$$

All the expectations are those of real-valued functions. We apply the ordinary result to each one and get (8.34). Thus, inducing up to  $n$  variables, conclude that if  $X_1, X_2, \dots, X_n$  are independent, then

$$E\left(\prod_1^n e^{iu_k X_k}\right) = \prod_1^n E e^{iu_k X_k}.$$

To go the other way, we make use of a result which will be proved in Chapter 11. If we consider the set of functions on  $R^{(n)}$ ,

$$\{e^{i(u_1 x_1 + \dots + u_n x_n)}\}, \quad \text{all } (u_1, \dots, u_n),$$

then these separate the  $n$ -dimensional distribution functions. Let  $F_n(\mathbf{x})$  be the distribution function of  $X_1, \dots, X_n$ ; then the left-hand side of the equation in proposition 8.33 is simply

$$\int \exp [i(u_1 x_1 + \dots + u_n x_n)] F_n(d\mathbf{x}).$$

But the right-hand side is the integral of

$$e^{i(u_1 x_1 + \dots + u_n x_n)}$$

with respect to the distribution function  $\prod_1^n F_k(x_k)$ . Hence  $F(x_1, \dots, x_n) = \prod_1^n F_k(x_k)$ , thus establishing independence.

**Notation.** To keep various variables and characteristic functions clear, we denote by  $f_X(u)$  the characteristic function of the random variable  $X$ .

**Corollary 8.35.** If  $X_1, \dots, X_n$  are independent, then the characteristic function of  $S_n = X_1 + \dots + X_n$  is given by

$$f_{S_n}(u) = \prod_1^n f_{X_k}(u).$$

The proof is obvious.

See that  $X_1, X_2$  independent implies that  $E e^{iu(X_1+X_2)} = f_{X_1}(u)f_{X_2}(u)$ . But having this hold for all  $u$  is not sufficient to guarantee that  $X_1, X_2$  are independent. (See Loève [108, p. 263, Example 1].)

Recall that in Chapter 3, we got the result that if  $X_1, X_2, \dots$  are independent,  $\sum_1^n X_k$  converges a.s. iff  $\sum_1^n X_k$  converges in probability, hence iff  $\sum_1^n X_k \xrightarrow{P} 0$ . The one obvious time that  $\xrightarrow{D}$  and  $\xrightarrow{P}$  coincide is when  $Y_n \xrightarrow{P} c \Leftrightarrow Y_n \xrightarrow{D} \delta(\text{degenerate at } c)$ . This observation will lead to

**Proposition 8.36.** For  $X_1, X_2, \dots$  independent,  $\sum_1^n X_k \xrightarrow{\text{a.s.}}$  iff  $\sum_1^n X_k \xrightarrow{D}$ , because for degenerate convergence, we can prove the following proposition.

**Proposition 8.37.** If  $Y_n$  are random variables with characteristic functions  $f_n(u)$ , then  $Y_n \xrightarrow{\mathcal{D}} 0$  iff  $f_n(u) \rightarrow 1$  in some neighborhood of the origin.

*Proof.* One way is obvious:  $Y_n \xrightarrow{\mathcal{D}} 0$  implies  $f_n(u) \rightarrow 1$  for all  $u$ . Now let  $f_n(u) \rightarrow 1$  in  $[-\delta, +\delta]$ . Proposition 8.29 gives

$$F_n\left(\left[-\frac{1}{\delta}, +\frac{1}{\delta}\right]^c\right) \leq \frac{\alpha}{\delta} \int_0^\delta (1 - \operatorname{Re} f_n(v)) dv.$$

The right-hand side goes to zero as  $n \rightarrow \infty$ , so the  $F_n$  are mass-preserving. Let  $n'$  be any subsequence such that  $F_{n'} \xrightarrow{\mathcal{D}} F$ . Then the characteristic function of  $F$  is identically one in  $[-\delta, +\delta]$ , hence  $F$  is degenerate at zero. By 8.7, the full sequence  $F_n$  converges to the law degenerate at zero.

This gives a criterion for convergence based on characteristic functions. Use the notation  $f_k(u) = f_{X_k}(u)$ .

**Theorem 8.38.**  $\sum_1^n X_k \xrightarrow{\text{a.s.}}$  iff  $\prod_1^\infty f_k(u)$  converges to  $h(u)$  in some neighborhood  $N$  of the origin, and  $|h(u)| > 0$  on  $N$ .

*Proof.* Certainly  $\sum_1^n X_k \xrightarrow{\text{a.s.}}$  implies  $\prod_1^\infty f_k(u)$  converges everywhere to a characteristic function. To go the other way, the characteristic function of  $\sum_1^n X_k$  is  $\prod_{k=1}^n f_k(u)$ . Because  $\prod_1^\infty f_k(u) \rightarrow h(u) \neq 0$  on  $N$ ,  $\prod_{k=1}^n f_k(u) \rightarrow 1$  on  $N$ . Use 8.37 to complete the proof, and note that 8.36 is a corollary.

## Problems

**22.** For  $Y_1, Y_2, \dots$  independent and  $\pm 1$  with probability  $\frac{1}{2}$ , use 8.38 to show that  $\sum c_k Y_k$  converges a.s.  $\Leftrightarrow \sum c_k^2 < \infty$ .

**23.** Show that the condition on  $f_k(u)$  in 8.38 can be partly replaced by—if  $\sum_1^\infty |1 - f_k(u)|$  converges in some neighborhood  $N$  of the origin, then  $\sum_1^n X_k \xrightarrow{\text{a.s.}}$ .

## 10. FOURIER INVERSION FORMULAS

To every characteristic function corresponds one and only one distribution function. Sometimes it is useful to know how, given a characteristic function, to find the corresponding distribution function, although by far the most important facts regarding characteristic functions do not depend on knowing how to perform this inversion. The basic inversion formula is the Fourier transform inversion formula. There are a lot of different versions of this; we give one particularly useful version.

**Theorem 8.39.** Let  $f(u)$  be the characteristic function of a distribution function  $F(dx)$  such that

$$\int |f(u)| du < \infty.$$

Then  $F(dx)$  has a bounded continuous density  $h(x)$  with respect to Lebesgue measure given by

$$(8.40) \quad h(x) = \frac{1}{2\pi} \int e^{-iux} f(u) du.$$

*Proof.* Assume that (8.40) holds true for one distribution function  $G(dx)$  with density  $g(x)$  and characteristic function  $\varphi(u)$ . Then we show that it holds true in general. Write

$$\frac{1}{2\pi} \int f(u) \varphi(u) e^{-iuv} du = \frac{1}{2\pi} \int e^{iux} e^{-iuv} \varphi(u) dF(x) du.$$

Then, interchanging order of integration on the right:

$$(8.41) \quad \frac{1}{2\pi} \int f(u) \varphi(u) e^{-iuv} du = \int g(y - x) dF(x)$$

If  $X$  has distribution  $F(dx)$ , and  $Y$  has distribution  $G(dx)$ , then the integral on the right is the density for the distribution of  $X + Y$  where they are taken to be independent. Instead of  $Y$ , now use  $\epsilon Y$ , in (8.41), because if the distribution of  $Y$  satisfies (8.40), you can easily verify that so does that of  $\epsilon Y$ , for  $\epsilon$  any real number. As  $\epsilon \rightarrow 0$  the characteristic function  $\varphi_\epsilon(u)$  of  $\epsilon Y$  converges to one everywhere. Use the bounded convergence theorem to conclude that the left-hand side of (8.41) converges to

$$h(y) = \frac{1}{2\pi} \int f(u) e^{-iuy} du.$$

The left-hand side is bounded by  $\int |f(u)| du$  for all  $y$ , so the integral of the left-hand side over any finite interval  $I$  converges to

$$\int_I h(y) dy.$$

If the endpoints of  $I$  are continuity points of  $F(x)$ , then since  $\mathfrak{L}(X + \epsilon Y) \rightarrow \mathfrak{L}(X)$ , the right-hand side of (8.41) converges to  $F(I)$ . Thus the two measures  $F(B)$  and  $\int_B h(y) dy$  on  $\mathcal{B}_1$  agree on all intervals, therefore are identical. The continuity and boundedness of  $h(x)$  follows directly from the expression (8.40). To conclude, all I have to do is produce one  $G(x)$ ,  $\varphi(u)$  for which (8.40) holds. A convenient pair is

$$(8.42) \quad g(x) = \frac{1}{2} e^{-|x|}, \quad \varphi(u) = \frac{1}{1 + u^2}.$$

To verify (8.42) do a straightforward contour integration.

We can use the same method to prove

**Proposition 8.43.** Let  $\varphi_n(u)$  be any sequence of characteristic functions converging to one for all  $u$  such that for each  $n$ ,

$$\int |\varphi_n(u)| du < \infty.$$

If  $b$  and  $a$  are continuity points of any distribution function  $F(x)$ , with characteristic function  $f(u)$ , then

$$F(b) - F(a) = \lim_n \frac{1}{2\pi} \int \left( \frac{e^{-iub} - e^{-iua}}{-iu} \right) \varphi_n(u) f(u) du$$

*Proof.* Whether or not  $F$  has a density or  $f(u)$  is integrable, (8.41) above still holds, where now the right-hand side is the density of the distribution of  $X + Y_n$ ,  $X, Y_n$  independent,  $\varphi_n(u)$  the characteristic function of  $Y_n$ . Since  $\varphi_n(u) \rightarrow 1, Y_n \xrightarrow{P} 0, \mathcal{L}(X + Y_n) \rightarrow \mathcal{L}(X)$ . The integral of the right-hand side over  $[a, b]$  thus converges to  $F(b) - F(a)$ . The integral of the left-hand side is

$$\frac{1}{2\pi} \int \left( \frac{e^{-iub} - e^{-iua}}{-iu} \right) \varphi_n(u) f(u) du.$$

This all becomes much simpler if  $X$  is distributed on the lattice  $L_d$ ,  $d > 0$ . Then

$$f(u) = \sum_{n=-\infty}^{+\infty} e^{iund} P(X = nd),$$

so that  $f(u)$  has period  $2\pi/d$ . The inversion formula is simply

$$P(X = nd) = \frac{d}{2\pi} \int_{-\pi/d}^{+\pi/d} e^{-iund} f(u) du$$

**Problem 24.** Let  $X_1, X_2, \dots$  be independent, identically distributed integer-valued random variables. Show that their sums are recurrent iff

$$\lim_{r \uparrow 1} \int_{-\pi}^{+\pi} \frac{1}{1 - rf(u)} du = \infty,$$

where  $f(u)$  is the common characteristic function of  $X_1, X_2, \dots$

## 11. MORE ON CHARACTERISTIC FUNCTIONS

There are some technical results concerning characteristic functions which we will need later. These revolve around expansions, approximation, and similar results.



**Proposition 8.44.** If  $E|X|^k < \infty$ , then the characteristic function of  $X$  has the expansion

$$f(u) = \sum_{j=0}^{k-1} \frac{(iu)^j}{j!} EX^j + \frac{(iu)^k}{k!} (EX^k + \delta(u)),$$

where  $\delta(u)$  denotes a function of  $u$ , such that  $\lim_{u \rightarrow 0} \delta(u) = 0$ , and satisfying  $|\delta(u)| \leq 3E|X|^k$  for all  $u$ .

*Proof.* Use the Taylor expansion with remainder on  $\sin y$ ,  $\cos y$  for  $y$  real to get

$$\cos y + i \sin y = \sum_0^{k-1} \frac{(iy)^j}{j!} + \frac{(iy)^k}{k!} (\cos \theta_1 y + i \sin \theta_2 y),$$

where  $\theta_1, \theta_2$  are real numbers such that  $|\theta_1| \leq 1, |\theta_2| \leq 1$ . Thus

$$e^{iuX} = \sum_0^k \frac{(iuX)^j}{j!} + \frac{(iuX)^k}{k!} [\cos(\theta_1 uX) + i \sin(\theta_2 uX) - 1].$$

Now  $\theta_1, \theta_2$  are random, but still  $|\theta_1| \leq 1, |\theta_2| \leq 1$ . Now

$$|X^k [\cos(\theta_1 uX) - 1 + i \sin(\theta_2 uX)]| \leq 3|X|^k,$$

which establishes  $|\delta(u)| \leq 3E|X|^k$ . Use the bounded convergence theorem to get

$$\lim_{u \rightarrow 0} \delta(u) = \lim_{u \rightarrow 0} E|X^k [\cos(\theta_1 uX) - 1 + i \sin(\theta_2 uX)]| = 0.$$

Another point that needs discussion is the logarithm of a complex number. For  $z$  complex,  $\log z$  is a many-valued function defined by

$$(8.45) \quad z = e^{\log z}.$$

For any determination of  $\log z$ ,  $\log z + 2n\pi i$ ,  $n = 0, \pm 1, \dots$  is another solution of (8.45). Write  $z = re^{i\theta}$ ; then  $\log z = \log r + i\theta$ . We always will pick that determination of  $\theta$  which satisfies  $-\pi < \theta \leq \pi$ , unless we state otherwise. With this convention,  $\log z$  is uniquely determined.

**Proposition 8.46.** For  $z$  complex,

$$\log(1+z) = z(1 + \epsilon(z)),$$

where  $|\epsilon(z)| \leq |z|$  for  $|z| \leq \frac{1}{2}$ .

*Proof.* For  $|z| < 1$ , the power series expansion is

$$\log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \dots = z \left( 1 - \frac{z}{2} + \frac{z^2}{3} - \dots \right).$$

For  $|z| \leq \frac{1}{2}$ ,

$$\left| -\frac{z}{2} + \frac{z^2}{3} - \dots \right| \leq \frac{|z|}{2} (1 + |z| + |z|^2 + \dots) = \frac{|z|}{2} \frac{1}{1 - |z|} \leq |z|.$$

One remark: Given a sequence  $f_n(u)$  of characteristic function, frequently we will take  $l_n(u) = \log f_n(u)$ , and show that  $l_n(u) \rightarrow \varphi(u)$  for some evaluation of the log function. Now  $l_n(u)$  is not uniquely determined.

$$l'_n(u) = l_n(u) + 2\pi i N_n(u),$$

$N_n(u)$  integer-valued, is just as good a version of  $\log f_n(u)$ . However, if  $l_n(u) \rightarrow \varphi(u)$ , and  $\varphi(u)$  is continuous at the origin for one evaluation of  $l_n(u)$ , then because  $f_n(u) = e^{l_n(u)} \rightarrow e^{\varphi(u)}$  the continuity theorem is in force.

## 12. METHOD OF MOMENTS

Suppose that all moments of a sequence of distribution functions  $F_n$  exists and for every integer  $k \geq 0$ , the limit of

$$\int x^k F_n(dx)$$

exists. Does it follow that there is a distribution  $F$  such that  $F_n \xrightarrow{\mathcal{D}} F$ ? Not necessarily! The reason that the answer may be "No" is that the functions  $x^k$  do not separate. There are examples [123] of distinct distribution functions  $F$  and  $G$  such that  $\int |x|^k dF < \infty$ ,  $\int |x|^k dG < \infty$  for all  $k \geq 0$ , and

$$\int x^k dF = \int x^k dG, \quad k = 0, 1, \dots$$

Start to argue this way: If  $\overline{\lim} \int x^2 dF_n < \infty$ , then (Problem 10) the  $\{F_n\}$  are mass-preserving. Take a subsequence  $F_{n'} \xrightarrow{\mathcal{D}} F$ . Then (Problem 14)

$$\lim_{n'} \int x^k dF_{n'} = \int x^k dF,$$

so for the full sequence

$$(8.47) \quad \lim_n \int x^k dF_n = \int x^k dF.$$

If there is only one  $F$  such that (8.47) holds, then every convergent subsequence of  $F_n$  converges to  $F$ , hence  $F_n \xrightarrow{\mathcal{D}} F$ . Thus

**Theorem 8.48.** *If there is at most one distribution function  $F$  such that*

$$\lim_n \int x^k dF_n = \int x^k dF,$$

*then  $F_n \xrightarrow{\mathcal{D}} F$ .*

The question is now one of uniqueness. Let

$$\mu_k = \lim_n \int x^k dF_n.$$

If  $F$  is uniquely determined by (8.47), then the moment problem given by the  $\mu_k$  is said to be determined. In general, if the  $\mu_k$  do not grow too fast, then uniqueness holds. A useful sufficient condition is

**Proposition 8.49.** *If*

$$\overline{\lim}_k \frac{|\mu_k|^{1/k}}{k} < \infty,$$

*then there is at most one distribution function  $F$  satisfying*

$$\mu_k = \int x^k F(dx).$$

*Proof.* Let

$$r = \overline{\lim}_k \frac{|\mu_k|^{1/k}}{k};$$

then for any  $\epsilon > 0$  and  $k \geq k_0$ , using the even moments to get bounds for the odd moments,

$$\int |x|^k dF \leq k^k(r + \epsilon)^k.$$

Hence, by the monotone convergence theorem,

$$\int e^{|\xi x|} dF = \lim_n \sum_{k=0}^n \frac{|\xi|^k}{k!} \int |x|^k dF < \infty,$$

for  $|\xi| < 1/re$ . Consider

$$\varphi(z) = \int e^{zx} F(dx).$$

By the above,  $\varphi(z)$  is analytic in the strip  $|R/z| < 1/re$ . For  $|z| < 1/re$ ,

$$(8.50) \quad \varphi(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!} \mu_k$$

This holds for any distribution function  $F$  having moments  $\mu_k$ . Since  $\varphi(z)$  in the strip is the analytic continuation of  $\varphi(z)$  given by (8.50), then  $\varphi(z)$  is completely determined by  $\mu_k$ . But for  $R/z = 0$ ,  $\varphi(z)$  is the characteristic function and thus uniquely determines  $F$ .

### 13. OTHER SEPARATING FUNCTION CLASSES

For restricted classes of distribution functions, there are separating classes of functions which are sometimes more useful than the complex exponentials. For example, consider only nonnegative random variables; their distribution functions assign zero mass to  $(-\infty, 0)$ . Call this class of distribution functions  $\mathcal{M}^+$ .

**Proposition 8.51.** *The exponentials  $\{e^{-\lambda x}\}$ ,  $\lambda$  real and nonnegative, separate in  $\mathcal{M}^+$ .*

*Proof.* Suppose  $F$  and  $G$  are in  $\mathcal{M}^+$  and for all  $\lambda \geq 0$ ,

$$\int e^{-\lambda x} F(dx) = \int e^{-\lambda x} G(dx).$$

Then substitute  $e^{-x} = y$ , so

$$(8.52) \quad \int_0^1 y^\lambda dF(-\log y) = \int_0^1 y^\lambda dG(-\log y).$$

In particular (8.52) holds for  $\lambda$  ranging through the nonnegative integers. Thus for any polynomial  $P(y)$ ,

$$\int_0^1 P(y) dF(-\log y) = \int_0^1 P(y) dG(-\log y),$$

hence equality holds for any continuous function on  $[0, 1]$ . Use an approximation argument to conclude now that  $F \equiv G$ .

As before, if  $F_n \in \mathcal{M}^+$  and  $\int e^{-\lambda x} F_n(dx)$  converges for all  $\lambda \geq 0$ , then there is at most one distribution function  $F$  such that  $F_n \xrightarrow{\mathcal{D}} F$ . Let the limit of  $\int e^{-\lambda x} dF_n(x)$  be  $h(\lambda)$ . Then by the bounded convergence theorem,

$$\lim_n \int \left( \frac{1 - e^{-\lambda x}}{\lambda x} \right) dF_n(x) = \frac{1}{\lambda} \int_0^\lambda h(\lambda) d\lambda.$$

So conclude, just as in the continuity theorem, that if

$$\lim_{\lambda \downarrow 0} h(\lambda) = 1,$$

then the sequence  $\{F_n\}$  is mass-preserving. Hence there is a unique distribution function  $F$  such that  $F_n \xrightarrow{\mathcal{D}} F$ .

For  $\mathbf{X}$  taking on nonnegative integer values, the *moment-generating* function is defined as

$$\varphi(z) = Ez^{\mathbf{X}},$$

for  $z$  complex,  $|z| \leq 1$ .

**Problem 25.** Prove that the functions  $z^x$ ,  $|z| \leq 1$  are separating in the class of distribution functions of nonnegative integer-valued random variables. If  $\{\mathbf{X}_n\}$  are a set of such random variables and

$$\varphi_n(z) = Ez_n^{\mathbf{X}_n}$$

converges for all  $|z| \leq 1$  to a function continuous at  $z = 1$ , then show there is a random variable  $\mathbf{X}$  such that  $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}$ .

## NOTES

More detailed background on distribution functions, etc., can be found in Loève's book [108]. For material on the moment problem, consult Shohat and Tamarkin [123]. For Laplace transforms of distributions  $\int e^{-\lambda x} dF(x)$  see Widder [140]. Although the central limit theorem for coin-tossing was proved early in the nineteenth century, a more general version was not formulated and proved until 1901 by Lyapunov [109]. The interesting proof we give in Section 6 is due to Lindeberg [106].

An important estimate for the rate of convergence in the central limit theorem is due to Berry and Eseen (see Loève, [108, pp. 282 ff.]). They prove that there is a universal constant  $c$  such that if  $S_n = X_1 + \cdots + X_n$  is a sum of independent, identically distributed random variables with  $EX_1 = 0$ ,  $EX_1^2 = \sigma^2 < \infty$ ,  $E|X_1|^3 < \infty$ , and if  $\Phi(x)$  is the distribution function of the  $N(0, 1)$  law, then

$$\sup_x \left| P\left(\frac{S_n}{\sigma\sqrt{n}} < x\right) - \Phi(x) \right| \leq c \frac{E|X_1|^3}{\sigma^3\sqrt{n}}.$$

It is known that  $c \leq 4$ , (Le Cam [99]) and unpublished calculations give bounds as low as 2.05. By considering coin-tossing, note that the  $1/\sqrt{n}$  rate of convergence cannot be improved upon.

## CHAPTER 9

# THE ONE-DIMENSIONAL CENTRAL LIMIT PROBLEM

### 1. INTRODUCTION

We know already that if  $X_1, X_2, \dots$  are independent and identically distributed,  $EX_1 = 0$ ,  $EX_1^2 = \sigma^2 < \infty$ , then

$$\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Furthermore, by the convergence of types theorem, no matter how  $S_n$  is normalized, if  $S_n/A_n \xrightarrow{\mathcal{D}}$  then the limit is a normal law or degenerate. So this problem is pretty well solved, with the exception of the question: Why is the normal law honored above all other laws? From here there are a number of directions available; the identically distributed requirement can be dropped. This leads again to a normal limit if some nice conditions on moments are satisfied. So the condition on moments can be dropped; take  $X_1, X_2, \dots$  independent, identically distributed but  $EX_1^2 = \infty$ . Now a new class of laws enters as the limits of  $S_n/A_n$  for suitable  $A_n$ , the so-called stable laws.

In a completely different direction is the law of rare events, convergence to a Poisson distribution. But this result is allied to the central limit problem and there is an elegant unification via the infinitely divisible laws. Throughout this chapter, unless explicitly stated otherwise, equations involving logs of characteristic functions are supposed to hold modulo additive multiples of  $2\pi i$ .

### 2. WHY NORMAL?

There is really no completely satisfying answer to this question. But consider, if  $X_1, X_2, \dots$ , are independent, identically distributed, and if

$$Z_n = \frac{X_1 + \dots + X_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} X$$

what are the properties that  $X$  must have? Look at

$$Z_{2n} = \frac{X_1 + \dots + X_n + X_{n+1} + \dots + X_{2n}}{\sqrt{2n}}.$$

Now  $Z_{2n} \xrightarrow{\mathcal{D}} X$ . But

$$Z_{2n} = \frac{X_1 + \cdots + X_n}{\sqrt{2}\sqrt{n}} + \frac{X_{n+1} + \cdots + X_{2n}}{\sqrt{2}\sqrt{n}} = \frac{1}{\sqrt{2}}(Z'_n + Z''_n).$$

The variables  $Z'_n, Z''_n$  are independent, and  $Z'_n \xrightarrow{\mathcal{D}} X, Z''_n \xrightarrow{\mathcal{D}} X$ , since they have the same distributions as  $Z_n$ . This (we would like to believe) implies that  $X$  has the same distribution as  $1/\sqrt{2}(X' + X'')$ , where  $X', X''$  are independent and have the same distribution as  $X$ . To verify this, note that

$$Ee^{iuZ_{2n}} = (Ee^{iuZ'_n/\sqrt{2}})(Ee^{iuZ''_n/\sqrt{2}})$$

or

$$f_{2n}(u) = f_n\left(\frac{u}{\sqrt{2}}\right)f_n\left(\frac{u}{\sqrt{2}}\right).$$

Since  $f_n(u) \xrightarrow{uc} f(u)$ , where  $f(u)$  is the characteristic function of  $X$ , it follows that  $f(u) = f(u/\sqrt{2})^2$ . But the right-hand side of this is the characteristic function of  $(X' + X'')/\sqrt{2}$ . So our expectation is fulfilled. Now the point is:

**Proposition 9.1.** *If a random variable  $X$  satisfies  $EX^2 < \infty$ , and*

$$X \stackrel{\mathcal{D}}{=} \frac{X' + X''}{\sqrt{2}},$$

*where  $X', X''$  are independent and  $\mathfrak{L}(X) = \mathfrak{L}(X') = \mathfrak{L}(X'')$ , then  $X$  has a  $\mathcal{N}(0, \sigma^2)$  distribution.*

*Proof.* The proof is simple. Let  $X_1, X_2, \dots$  be independent,  $\mathfrak{L}(X_k) = \mathfrak{L}(X)$ .  $EX$  must be zero, since  $EX = (EX' + EX'')/\sqrt{2}$  implies  $EX = \sqrt{2}EX$ . By iteration,

$$X \stackrel{\mathcal{D}}{=} \frac{X_1 + \cdots + X_n}{\sqrt{n}}, \quad n = 2^m.$$

But the right-hand sums, divided by  $\sigma$ , converge in distribution to  $\mathcal{N}(0, 1)$ .

Actually, this result holds without the restriction that  $EX^2 < \infty$ . A direct proof of this is not difficult, but it also comes out of later work we will do with stable laws, so we defer it.

### 3. THE NONIDENTICALLY DISTRIBUTED CASE

Let  $X_1, X_2, \dots$  be independent. Then

**Theorem 9.2.** *Let  $EX_k = 0, EX_k^2 = \sigma_k^2 < \infty, E|X_k|^3 < \infty$ , and  $s_n^2 = \sum_1^n \sigma_k^2$ . If*

$$(9.3) \quad \lim_n \frac{1}{s_n^3} \sum_1^n E|X_k|^3 = 0,$$

then

$$\frac{X_1 + \cdots + X_n}{s_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

*Proof.* Very straightforward and humdrum, using characteristic functions. Let  $f_k$  be  $f_{X_k}$ ,  $g_n$  the characteristic function of  $S_n/s_n$ . Then

$$g_n(u) = \prod_1^n f_k\left(\frac{u}{s_n}\right) = \prod_1^n \left[1 + \left(f_k\left(\frac{u}{s_n}\right) - 1\right)\right].$$

Using the Taylor expansion, we get from (8.44)

$$\begin{aligned} f_k\left(\frac{u}{s_n}\right) &= 1 - \frac{u^2}{2s_n^2} \left(\sigma_k^2 + \delta_k\left(\frac{u}{s_n}\right)\right), \\ \left|f_k\left(\frac{u}{s_n}\right) - 1\right| &\leq 2u^2 \frac{\sigma_k^2}{s_n^2}. \end{aligned}$$

Now  $(E|X_k^3|)^{3/2} < E|X_k|^3$ , or  $\sigma_k^3 \leq E|X_k|^3$ . Then condition 9.3 implies that  $\sup_{k \leq n} \sigma_k/s_n \rightarrow 0$ . So  $\sup_{k \leq n} |f_k(u/s_n) - 1| \rightarrow 0$  as  $n$  goes to infinity. Therefore use the log expansion

$$\log(1 + z) = z(1 + \theta z),$$

where  $|\theta| \leq 1$  for  $|z| \leq \frac{1}{2}$ , to get

$$\log g_n(u) = \sum_1^n \left(f_k\left(\frac{u}{s_n}\right) - 1\right) + \theta \sum_1^n \left(f_k\left(\frac{u}{s_n}\right) - 1\right)^2,$$

where the equality holds modulo  $2\pi i$ . Consider the second term above,

$$\begin{aligned} \left|\theta \sum_1^n \left(f_k\left(\frac{u}{s_n}\right) - 1\right)^2\right| &\leq \sup_{k \leq n} \left|f_k\left(\frac{u}{s_n}\right) - 1\right| \cdot \sum_1^n \left|f_k\left(\frac{u}{s_n}\right) - 1\right| \\ &\leq \sup_{k \leq n} \left|f_k\left(\frac{u}{s_n}\right) - 1\right| \cdot \sum_1^n 2u^2 \frac{\sigma_k^2}{s_n^2} \\ &\leq 2u^2 \sup_{k \leq n} \left|f_k\left(\frac{u}{s_n}\right) - 1\right|. \end{aligned}$$

This bound goes to zero as  $n \rightarrow \infty$ . Apply the Taylor expansion,

$$f_k\left(\frac{u}{s_n}\right) - 1 = -\frac{u^2}{2} \frac{\sigma_k^2}{s_n^2} + \theta_k \frac{u^3}{s_n^3} E|X_k|^3, \quad |\theta_k| \leq 1,$$

to the first term above to get

$$\sum_1^n \left(f_k\left(\frac{u}{s_n}\right) - 1\right) = -\frac{u^2}{2} + \frac{u^3 \theta}{s_n^3} \sum_1^n E|X_k|^3, \quad |\theta| \leq 1,$$

which converges to  $-u^2/2$ .



We conclude that

$$g_n(n) \rightarrow e^{-u^2/2}$$

for every  $u$ . Since the theorem holds for identically distributed random variables, it follows that  $e^{-u^2/2}$  must be the characteristic function of the  $\mathcal{N}(0, 1)$  distribution. Apply the continuity theorem to complete the proof.

Note that we got, in this proof, the additional dividend that if  $X$  is  $\mathcal{N}(0, 1)$ , then

$$Ee^{iuX} = e^{-u^2/2}.$$

#### 4. THE POISSON CONVERGENCE

For  $X_1, X_2, \dots$  independent and identically distributed,  $EX_1 = 0$ ,  $EX_1^2 = \sigma^2$ , let

$$M_n = \max \left( \frac{|X_1|}{\sqrt{n}}, \dots, \frac{|X_n|}{\sqrt{n}} \right).$$

Write, for  $x > 0$ ,

$$\begin{aligned} P(M_n < x) &= P(|X_1| < \sqrt{nx}, \dots, |X_n| < \sqrt{nx}) \\ &= [P(|X_1| < \sqrt{nx})]^n. \end{aligned}$$

Now

$$P(|X_1| \geq \sqrt{nx}) \leq \frac{1}{nx^2} \int_{|y| \geq \sqrt{nx}} y^2 dF(y) = \frac{1}{n} \theta_n(x),$$

where  $\lim_n \theta_n(x) = 0$ . This leads to

$$P(M_n < x) \geq \left( 1 - \frac{1}{n} \theta_n(x) \right)^n \approx e^{-\theta_n(x)},$$

the point being that

$$\lim_n P(M_n < x) = 1, \quad \text{all } x,$$

or  $M_n \xrightarrow{\mathcal{D}} 0$ . In this case, therefore, we are dealing with sums

$$Z_n = \left( \frac{X_1}{\sqrt{n}} \right) + \dots + \left( \frac{X_n}{\sqrt{n}} \right) = X_1^{(n)} + \dots + X_n^{(n)}$$

of independent random variables such that the maximum of the individual summands converges to zero. I have gone through this to contrast it to the situation in which we have a sequence of coins, 1, 2,  $\dots$ , with probabilities of heads  $p_1, p_2, \dots$ , where  $p_n \rightarrow 0$ , and the  $n$ th coin is tossed  $n$  times. For the  $n$ th coin, let  $X_k^{(n)}$  be one if heads comes on the  $k$ th trial, zero otherwise. So

$$S_n = X_1^{(n)} + \dots + X_n^{(n)}$$

is the number of heads gotten using the  $n$ th coin. Think!—the probability of heads on each individual trial is  $p_n$  and that is going to zero. However, the total number of trials is getting larger and larger. Is it possible that  $S_n$  converges in distribution? Compute

$$P(S_n = 0) = (1 - p_n)^n.$$

This will converge if and only if  $np_n \rightarrow \lambda$ ,  $0 \leq \lambda < \infty$ . If  $\lambda = 0$ , then  $P(S_n = 0) \rightarrow 1$ , and henceforth rule this case out.

For  $\lambda > 0$ ,  $P(M_n = 0) = P(S_n = 0) \rightarrow e^{-\lambda}$ , so  $M_n \xrightarrow{\mathcal{D}} 0$ . Take characteristic functions, noting that

$$Ee^{iuX_k^{(n)}} = p_n e^{iu} + q_n.$$

For  $n$  sufficiently large, these are close to one, for  $u$  fixed, and we can write

$$f_{S_n}(u) = (1 + p_n(e^{iu} - 1))^n,$$

$$\log f_{S_n}(u) = np_n(e^{iu} - 1) + \theta np_n^2(e^{iu} - 1)^2, \quad |\theta| \leq 1.$$

Since  $p_n \rightarrow 0$ ,  $np_n \rightarrow \lambda$ ,  $np_n^2 \rightarrow 0$ , this gives

$$\log f_{S_n}(u) \rightarrow \lambda(e^{iu} - 1),$$

so

**Theorem 9.4.**  $S_n \xrightarrow{\mathcal{D}}$  if and only if  $np_n \rightarrow \lambda$ , then the limit has characteristic function  $\exp[\lambda(e^{iu} - 1)]$ . The limit random variable  $X$  takes values in  $\{0, 1, 2, \dots\}$ , so

$$Ee^{iuX} = \sum_0^\infty e^{iuk} P(X = k).$$

Expanding,

$$e^{\lambda(e^{iu} - 1)} = \sum_0^\infty e^{-\lambda} \cdot \frac{\lambda^k}{k!} e^{iuk},$$

so

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

**Definition 9.5.** A random variable  $X$  taking values in  $\{0, a, 2a, 3a, \dots\}$  will be said to have Poisson distribution  $\mathcal{P}(\lambda)$ ,  $\lambda \geq 0$  with jump size  $a$ , if

$$P(X = ak) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{or} \quad f_X(u) = \exp[\lambda(e^{iua} - 1)].$$

Look now at the  $X_1^{(n)}, \dots, X_n^{(n)}$ . Since  $P(X_k^{(n)} = 0) = q_n \rightarrow 1$ , usually the  $X_k^{(n)}$  are zero, but once in a while along comes a blip. Again, take  $M_n = \max(X_1^{(n)}, \dots, X_n^{(n)})$ . Now  $M_n$  can only take the values 0 or 1, and  $M_n \rightarrow 0$  unless  $\lambda = 0$ . Here the contrast obviously is that  $M_n$  must equal 1 with

positive probability, or  $S_n \xrightarrow{\mathcal{D}} 0$ . It is the difference between the sum of uniformly small smears, versus the sum of occasionally large blips. That this is pretty characteristic is emphasized by

**Proposition 9.6.** Let  $S_n = X_1^{(n)} + \cdots + X_n^{(n)}$ , where  $X_1^{(n)}, \dots, X_n^{(n)}$  are independent and identically distributed. If  $S_n \xrightarrow{\mathcal{D}} X$  then  $M_n \xrightarrow{\mathcal{D}} 0$  if and only if  $X$  is normal.

*Proof.* Deferred until Section 7.

The Poisson convergence can be generalized enormously. For example, suppose  $S_n = X_1^{(n)} + \cdots + X_n^{(n)}$ , the  $X_k^{(n)}$  independent and identically distributed with

$$X_k^{(n)} = \begin{cases} 0, & \text{with probability } q_n \\ x_i, & \text{with probability } p_n^{(i)}, \quad i = 1, \dots, j. \end{cases}$$

and  $S_n \xrightarrow{\mathcal{D}} X$ . We could again show that this is possible only if  $np_n^{(i)} \rightarrow \lambda_i$ ,  $0 \leq \lambda_i < \infty$ , and if so, then

$$f_X(u) = \exp \left( \sum_{i=1}^j \lambda_i (e^{iux_i} - 1) \right).$$

Two interesting points are revealed in this result.

*First:* The expected number of times that  $X_k^{(n)} = x_i$  is  $np_n^{(i)}$ . So  $\lambda_i$  is roughly the expected number of times that one of the summands is  $x_i$ .

*Second:* Since

$$f_X(u) = \prod_{i=1}^j \exp(\lambda_i (e^{iux_i} - 1)),$$

$X$  is distributed as  $X_1 + \cdots + X_j$ , where the  $X_i$  are independent random variables and  $X_i$  has Poisson distribution  $\mathcal{P}(\lambda_i)$  with jump size  $x_i$ . So the jumps do not interact; each jump size  $x_i$  contributes an independent Poisson component.

## 5. THE INFINITELY DIVISIBLE LAWS

To include both Poisson and  $\mathcal{N}(0, 1)$  convergence, ask the following question: Let

$$(9.7) \quad S_n = X_1^{(n)} + \cdots + X_n^{(n)},$$

where the  $X_k^{(n)}$ ,  $k = 1, \dots, n$ , are independent and identically distributed.

If  $S_n \xrightarrow{\mathcal{D}} X$ , what are the possible distributions of  $X$ ?

$S_n$  is the sum of many independent components; heuristically  $X$  must have this same property.

**Definition 9.8.**  $X$  will be said to have an infinitely divisible distribution if for every  $n$ , there are independent and identically distributed random variables  $X_1^{(n)}, \dots, X_n^{(n)}$  such that  $\mathbb{L}(X) = \mathbb{L}(X_1^{(n)} + \dots + X_n^{(n)})$ .

**Proposition 9.9.** A random variable  $X$  is a limit in distribution of sums of the type (9.7) if and only if it has an infinitely divisible distribution.

*Proof.* If  $X$  has an infinitely divisible distribution, then by definition there are sums  $S_n$  of type (9.7) with distribution exactly equal to  $X$ . The other way: Consider

$$S_{2n} = (X_1^{(2n)} + \dots + X_n^{(2n)}) + (X_{n+1}^{(2n)} + \dots + X_{2n}^{(2n)}) = Y_n + Y'_n.$$

The random variables  $Y_n$  and  $Y'_n$  are independent with the same distribution. If  $S_n \xrightarrow{\mathcal{D}} X$ , the distributions of  $Y_n$  are mass-preserving, because

$$(P(Y_n > y))^2 = P(Y_n > y, Y'_n > y) \leq P(S_{2n} > 2y),$$

and similarly,

$$(P(Y_n < -y))^2 \leq P(S_{2n} < -2y).$$

Take a subsequence  $\{n'\}$  such that  $Y_{n'} \xrightarrow{\mathcal{D}} Y$ . Obviously,  $f_X(u) = [f_Y(u)]^2$ ; so  $\mathbb{L}(X) = \mathbb{L}(Y + Y')$ ,  $Y, Y'$  independent. This can be repeated to get  $X$  equal in distribution to the sum of  $Y_1^{(n)} + \dots + Y_n^{(n)}$  by considering  $S_{nm}$ .

If  $S_n \xrightarrow{\mathcal{D}} X$ , do the components  $X_1^{(n)}, \dots, X_n^{(n)}$  have to get smaller and smaller in any reasonably formulated way? Note that in both the Poisson and  $\mathcal{N}(0, 1)$  convergence, for any  $\epsilon > 0$ ,  $P(|X_1^{(n)}| \geq \epsilon) \rightarrow 0$ ; that is,  $X_1^{(n)} \xrightarrow{\mathcal{D}} 0$  [so, of course,  $\sup_k P(|X_k^{(n)}| \geq \epsilon) \rightarrow 0$ , since these probabilities are the same for all  $k = 1, \dots, n$ ]. This holds in general.

**Proposition 9.10.** If  $S_n \xrightarrow{\mathcal{D}} X$ , then  $X_1^{(n)} \xrightarrow{\mathcal{D}} 0$ .

*Proof.* Since  $f_{S_n}(u) \xrightarrow{uc} f_X(u)$ , there is a neighborhood  $N$  of the origin such that  $\operatorname{Re} f_{S_n}(u) \geq \delta > 0$ ,  $u \in N$ , all  $n \geq 1$ . On  $N$ ,  $|\operatorname{Arg} f_{S_n}(u)|$  is bounded away from  $\pi$ . Let  $f_n(u) = f_{X_1^{(n)}}(u)$ ; so

$$f_{S_n}(u) = [f_n(u)]^n.$$

On  $N$ , then

$$|f_{S_n}(u)| = |f_n(u)|^n,$$

$$\operatorname{Arg} f_{S_n}(u) = n \operatorname{Arg} f_n(u).$$

So  $f_n(u) \rightarrow 1$ , for  $u \in N$ , and now apply 8.37.

Now I turn to the problem of characterizing the infinitely divisible distributions. Let  $f(u)$  be the characteristic function of  $X$ . Therefore, since

$\mathfrak{L}(X) = \mathfrak{L}(X_1^{(n)} + \cdots + X_n^{(n)})$ , there is a characteristic function  $f_n(u)$  such that  $f(u) = [f_n(u)]^n$ , and by 9.10,  $f_n(u) \xrightarrow{uc} 1$ . Then,

$$\begin{aligned} \log f(u) &= n \log [1 - (1 - f_n(u))] \\ (9.11) \quad &= n(f_n(u) - 1)(1 + \epsilon_n(u)). \end{aligned}$$

Since  $f_n(u) \xrightarrow{uc} 1$ , it follows that  $\epsilon_n(u) \xrightarrow{uc} 0$ . Also,  $|f(u)| > 0$ , all  $u$ , otherwise  $f(u_0) = 0$  implies  $f_n(u_0) = 0$ , all  $n$ , contradicting  $f_n(u) \xrightarrow{uc} 1$ . Denote by  $F_n$  the distribution function of  $X_1^{(n)}$ ; then

$$(9.12) \quad \log f(u) = (1 + \epsilon_n(u)) \int (e^{iuz} - 1) n dF_n.$$

If we set up approximating sums of the integral in (9.12), we get

$$f(u) = \exp \left[ \sum \lambda_i (e^{iuz_i} - 1) \right],$$

exactly like the general Poisson case looked at before. Note also that if we put  $\mu_n(B) = nF_n(B)$ ,  $\mu_n$  a nonnegative measure on  $\mathcal{B}_1$ , then

$$\mu_n(B) = E(\text{number of } X_k^{(n)} \in B).$$

Since  $\log f(u) = (1 + \epsilon_n) \int (e^{iuz} - 1) \mu_n(dx)$ , if  $\mu_n$  converges to a measure  $\mu$  such that for continuous bounded functions  $\varphi(x)$ ,  $\int \varphi(x) \mu_n(dx) \rightarrow \int \varphi(x) \mu(dx)$ , then we could conclude

$$f(u) = e^{\int (e^{iuz} - 1) \mu(dx)}.$$

This is the basic idea, but there are two related problems. First, the total mass of  $\mu_n$  is  $\mu_n(R^{(1)}) = n$ , hence  $\mu_n(R^{(1)}) \rightarrow \infty$ . Certainly, then, for  $\varphi(x) = 1$  there is no finite  $\mu$  such that  $\int \varphi d\mu_n \rightarrow \int \varphi d\mu$ . Second, how can the  $\mathcal{N}(0, 1)$  characteristic function  $e^{-u^2/2}$  be represented as above? Now, for any neighborhood  $N$  of the origin, we would expect more and more of the  $X_k^{(n)}$  to be in  $N$ ; that is,  $\mu_n(N) \rightarrow \infty$ . But in analogy with Poisson convergence, the number of times that  $X_k^{(n)}$  is sizeable enough to take values outside of  $N$  should be bounded; that is,  $\overline{\lim} \mu_n(N^c) < \infty$ . We can prove even more than this.

**Proposition 9.13**

$$\overline{\lim} \mu_n[-a, +a]^c \leq \alpha a \int_0^{1/a} |Rl \log f(v)| dv.$$

*Proof.* By inequality 8.29,

$$\mu_n[-a, +a]^c \leq \alpha n \int_0^{1/a} (1 - Rl f_n(v)) dv.$$

Take the real part of Eq. (9.11), and pass to the limit to get

$$\lim_n n(1 - Rlf_n(u)) = -Rl \log f(u).$$

Use  $|f(u)| > 0$ , all  $u$ , and the bounded convergence theorem for the rest.

Since

$$Rl \log f(0) = 0, \quad \overline{\lim}_{a \uparrow \infty} \overline{\lim}_n \mu_n[-a, +a]^c = 0,$$

so the  $\mu_n$  sequence is in this sense mass-preserving. What is happening is that the mass of  $\mu_n$  is accumulating near the origin, and behaving nicely away from the origin as  $n \rightarrow \infty$ . But if then  $\varphi(0) = 0$ , there is some hope that  $\int \varphi(x) \mu_n(dx)$  may converge. This is true to some extent, more exactly,

**Proposition 9.14**

$$\overline{\lim} \int_{[-1, +1]} x^2 \mu_n(dx) < \infty.$$

*Proof.*

$$\begin{aligned} n(1 - Rlf_n(1)) &= \int (1 - \cos x) \mu_n(dx) \\ &\geq \int_{[-1, +1]} (1 - \cos x) \mu_n(dx). \end{aligned}$$

By Taylor's expansion,  $\cos x = 1 - x^2/2 \cos x\alpha$ ,  $|\alpha| \leq 1$ , so there is a  $\beta$ ,  $0 < \beta < \infty$ , such that  $\cos x \leq 1 - \beta x^2$ , for  $|x| \leq 1$ . Thus

$$n(1 - Rlf_n(1)) \geq \beta \int_{[-1, +1]} x^2 \mu_n(dx).$$

However,  $n(1 - Rlf_n(1)) \rightarrow Rl \log f(1)$ , giving the result, since  $|f(1)| \neq 0$ .

By 9.13 and 9.14, if we define  $\nu_n(B) = \int_B \varphi(x) \mu_n(dx)$ , where  $\varphi(x)$  is bounded and behaves like  $x^2$  near the origin, the  $\nu_n(B)$  is a bounded sequence of measures and we can think of trying to apply the Helly-Bray theorem. The choice of  $\varphi(x)$  is arbitrary, subject only to boundedness and the right behavior near zero. The time honored custom is to take  $\varphi(x)$  to be  $x^2/(1 + x^2)$ . Thus let  $\alpha_n = \int (y^2/(1 + y^2)) \mu_n(dy)$ ,

$$G_n(x) = \frac{1}{\alpha_n} \int_{(-\infty, x)} \left( \frac{y^2}{1 + y^2} \right) \mu_n(dy),$$

making  $G_n(x)$  a distribution function. By 9.13 and 9.14  $\overline{\lim} \alpha_n < \infty$ . We can write

$$\log f(u) = (1 + \epsilon_n) \alpha_n \int (e^{iux} - 1) \frac{x^2 + 1}{x^2} G_n(dx),$$

but the integrand blows up as  $x \rightarrow 0$ . So we first subtract the infinity by writing,

$$\log f(u) = (1 + \epsilon_n) \int \left( e^{iux} - 1 - \frac{iux}{1 + x^2} \right) d\mu_n + (1 + \epsilon_n) \int \frac{iux}{1 + x^2} d\mu_n.$$

Then we write  $\beta_n = \int x/(1+x^2) d\mu_n$ , so that

$$(9.15) \quad \log f(u) = (1 + \epsilon_n)\alpha_n \int \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{1+x^2}{x^2} G_n(dx) \\ + i(1 + \epsilon_n)\beta_n u.$$

If the integral term converges we see that  $\{\beta_n\}$  can contain no subsequence going to infinity. If it did, then

$$\exp[i\beta_n u(1 + \epsilon_n(u))] \xrightarrow{uc} h(u)$$

would imply that on substituting  $u = v/\beta_n$  and going along the subsequence, we would get  $e^{iv} = h(0)$  for all  $v$ . If  $\{\beta_n\}$  has two limit points  $\beta, \beta'$ , then  $e^{i\beta u} = e^{i\beta' u}$ , hence  $\beta = \beta'$ . Thus,  $uc$  convergence of the first term entails convergence of the second term to  $i\beta u$ .

The integrand in (9.15)

$$\varphi(x, u) = \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{1+x^2}{x^2}$$

is a continuous bounded function of  $x$  for  $x \neq 0$ . As  $x \rightarrow 0$ , it has the limit  $-u^2/2$ . By defining  $\varphi(0, u) = -u^2/2$ ,  $\varphi(x, u)$  is jointly continuous everywhere. By 9.13,  $\{G_n\}$  is mass-preserving. If  $\lim \alpha_n = 0$ , take  $n'$  such that  $\alpha_{n'} \rightarrow 0$  and conclude from (9.15) that  $X$  is degenerate. Otherwise, take  $n'$  such that  $\alpha_{n'} \rightarrow \alpha > 0$  and  $G_{n'} \xrightarrow{D} G$ . Then  $G$  is a distribution function. Go along the  $n'$  sequence in 9.15. The  $\beta_{n'}$  sequence must converge to some limit  $\beta$  since the integral term converges  $uc$ . Therefore

$$\log f(u) = \alpha \int \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{1+x^2}{x^2} G(dx) + i\beta u.$$

Suppose  $G(\{0\}) > 0$ , then

$$(9.16) \quad \log f(u) = i\beta u - \frac{\sigma^2 u^2}{2} + \alpha \int_{\{0\}^c} \varphi(x, u) G(dx).$$

We have now shown part of

**Theorem 9.17.**  $X$  has infinitely divisible distribution if and only if its characteristic function  $f(u)$  is given by

$$(9.18) \quad \log f(u) = i\beta u - \frac{\sigma^2 u^2}{2} + \int \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{1+x^2}{x^2} \nu(dx),$$

where  $\nu$  is a finite measure that assigns zero mass to the origin.

*To complete the proof:* It has to be shown that any random variable whose characteristic function is of the form (9.18) has infinitely divisible

distribution. To begin, assume that any function  $f(u)$  whose log is of the form (9.18) is a characteristic function. Then it is trivial because if  $f_n(u)$  is defined by

$$\log f_n(u) = \frac{1}{n} \log f(u),$$

then  $\log f_n(u)$  is again of the form (9.18); so  $f_n(u)$  is a characteristic function. Since  $f(u)$  now is given by  $[f_n(u)]^n$  for any  $n$ , the corresponding distribution is infinitely divisible. The last point is to show that (9.18) always gives a characteristic function. Take partitions  $\mathcal{I}_n$  of  $R^{(1)}$  into finite numbers of intervals such that the Riemann sums in (9.18) converge to the integral, that is,

$$\sum_{i=1}^{j_n} \left( e^{iux_i} - 1 - \frac{iux_i}{1+x_i^2} \right) \frac{1+x_i^2}{x_i^2} \nu(I_i) \rightarrow \int \varphi(x, u) \nu(dx).$$

Put  $\beta_n = \beta - \sum_i (\nu(I_i)/x_i)$ , denote

$$\lambda_i = \nu(I_i) \cdot \frac{1+x_i^2}{x_i^2},$$

and write

$$g_n(u) = (\exp [i\beta_n u - (\sigma^2/2)u^2]) \cdot (\exp [\sum_i \lambda_i (e^{iux_i} - 1)]).$$

See that  $g_n(u)$  is the product of a characteristic function of a  $\mathcal{N}(\beta_n, \sigma)$  distribution and characteristic functions of Poisson distributions  $\mathcal{P}(\lambda_i)$  with jump  $x_i$ . Therefore by Corollary 8.35,  $g_n(u)$  is a characteristic function. This does it, because

$$\lim_n g_n(u) = f(u)$$

for every  $u$ . Check that anything of the form (9.18) is continuous at  $u = 0$ . Certainly the first two terms are. As to the integral, note that  $\sup_x |\varphi(x, u)| \leq M$  for all  $|u| \leq 1$ . Also,  $\lim_{u \rightarrow 0} \varphi(x, u) = 0$  for every  $x$ , and apply the bounded convergence theorem to get

$$\lim_{u \rightarrow 0} \int \varphi(x, u) \nu(dx) = 0, \quad \text{so} \quad \lim_{u \rightarrow 0} f(u) = 1.$$

By the continuity theorem,  $f(u)$  is a characteristic function. Q.E.D.

## 6. THE GENERALIZED LIMIT PROBLEM

Just as before, it becomes reasonable to ask what are the possible limit laws of

$$S_n = X_1^{(n)} + \cdots + X_n^{(n)}$$

if the restriction that the  $X_1^{(n)}, \dots, X_n^{(n)}$  be identically distributed is lifted.



Some restriction is needed; otherwise, take

$$X_2^{(n)} = \dots = X_n^{(n)} = 0, \quad X_1^{(n)} = X,$$

to get any limit distribution desired. What is violated in the spirit of the previous work is the idea of a sum of a large number of components, each one *small on the average*. That is, we had, in the identically distributed case, that for every  $\epsilon > 0$ ,

$$(A) \quad \sup_k P(|X_k^{(n)}| > \epsilon) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

This is the restriction that we retain in lieu of identical distribution. It is just about the weakest condition that can be imposed on the summands in order to prevent one of the components from exerting a dominant influence on the sum. With condition (A) a surprising result comes up.

**Theorem 9.19.** *If the sums  $S_n \xrightarrow{\mathcal{D}} X$ , then  $X$  has infinitely divisible distribution.*

So in a strong sense the infinitely divisible laws are the limit laws of large sums of independent components, each one small on the average. The proof of 9.19 proceeds in exactly the same way as that of Theorem 9.17, the only difference being that  $\mu_n(B) = \sum_1^n F_k^{(n)}(B)$  instead of  $nF_n(B)$ , but the same inequalities are used. It is the same proof except that one more subscript is floating around.

**Problem 1.** Let  $X$  have infinitely divisible distribution,

$$\log f(u) = i\beta u + \int \varphi(x, u)\nu(dx)$$

if  $\nu(\{0\}) = 0$ , and if  $\nu$  assigns all its mass to a countable set of points, prove that the distribution of  $X$  is of pure type. [Use the law of pure types.]

## 7. UNIQUENESS OF REPRESENTATION AND CONVERGENCE

Let  $X$  have an infinitely divisible distribution with characteristic function  $f(u)$ . Then by (9.18), there is a finite measure  $\gamma(dx)$  (possibly with mass at the origin) and a constant  $\beta$  such that

$$(9.20) \quad \log f(u) = i\beta u + \int \varphi(x, u)\gamma(dx) \quad [2\pi i],$$

and  $\varphi(x, u)$  is continuous in both  $x$  and  $u$  and bounded for  $x \in R^{(1)}$ ,  $u \in [-U, +U]$ ,  $U < \infty$ .  $\log f(u)$  is defined up to additive multiples of  $2\pi i$ . Because  $|f(u)| \neq 0$ , there is a unique version of  $\log f(u)$  which is zero when  $u$  is zero and is a continuous function of  $u$  on  $R^{(1)}$ . Now (9.20) states that this version is given by the right-hand side above.

**Proposition 9.21.**  $\gamma(dx)$  and  $\beta$  are uniquely determined by (9.20).

*Proof.* Let  $\psi(u) = \log f(u)$ ; this is the continuous version of  $\log f(u)$ . Then (following Loève [108]), take

$$\theta(u) = \psi(u) - \int_0^1 \left[ \frac{\psi(u+h) + \psi(u-h)}{2} \right] dh$$

so that  $\theta(u)$  is determined by  $\psi$ , hence by  $f$ . Note that

$$\begin{aligned} \int_0^1 \left[ \frac{\varphi(x, u+h) + \varphi(x, u-h)}{2} \right] dh \\ = \varphi(x, u) - e^{iux} \int_0^1 (1 - \cos(hx)) dh \cdot \frac{1+x^2}{x^2} \\ = \varphi(x, u) - e^{iux} \left( 1 - \frac{\sin x}{x} \right) \cdot \frac{1+x^2}{x^2}. \end{aligned}$$

Hence, using (9.20),

$$\theta(u) = \int e^{iux} g(x) \gamma(dx),$$

where

$$g(x) = \left( 1 - \frac{\sin x}{x} \right) \frac{1+x^2}{x^2}.$$

It is easy to check that  $0 < \inf g(x) < \sup g(x) < \infty$ . But  $\theta(u)$  uniquely determines the measure  $\nu(B) = \int_B g(x) \gamma(dx)$ , and thus  $\gamma$  is determined as  $\gamma(B) = \int_B [g(x)]^{-1} \nu(dx)$ . If, therefore,

$$\psi(u) = i\beta'u + \int \varphi(x, u) \gamma'(dx),$$

then  $\gamma \equiv \gamma'$  implying  $\beta = \beta'$ .

The fact that  $\gamma(dx)$  is unique gives us a handhold on conditions for  $S_n \xrightarrow{\mathcal{D}} X$ . Let  $\gamma(dx) = \alpha G(dx)$  where  $G$  is a distribution function. Recall that  $\alpha$ ,  $G(x)$  were determined by taking any convergent subsequence  $\alpha_{n'}$  of  $\alpha_n$ ,  $\alpha = \lim_{n'} \alpha_{n'}$ , and taking  $G(x)$  as any limit distribution of the  $G_{n'}(x)$  sequence. Since  $\alpha$ ,  $G(x)$  are unique, then  $\alpha_n \rightarrow \alpha$ ,  $G_n \xrightarrow{\mathcal{D}} G$ . Consequently  $\beta_n$ , defined by

$$\beta_n = \int \frac{x}{1+x^2} d\mu_n,$$

converges to  $\beta$ . Thus, letting  $\gamma(x) = \gamma(-\infty, x)$ , and

$$\gamma_n(x) = \int_{(-\infty, x)} \frac{y^2}{1+y^2} \mu_n(dy),$$

then  $S_n \xrightarrow{\mathcal{D}} X$  implies  $\gamma_n \xrightarrow{\mathcal{D}} \gamma$ , and  $\{\gamma_n\}$  is mass-preserving in the sense that for any  $\epsilon > 0$ , there exists a finite interval  $I$  such that  $\sup_n \gamma_n(I^c) \leq \epsilon$ . These conditions are also sufficient:

**Theorem 9.22.**  $S_n \xrightarrow{\mathcal{D}} X$  where  $X$  has characteristic function given by

$$\log f(u) = i\mu\beta + \int \varphi(x, u)\gamma(dx)$$

for  $\gamma(dx)$  a finite measure if and only if the measures  $\gamma_n(dx)$  are mass-preserving in the above sense and

$$\text{i) } \gamma_n \xrightarrow{\mathcal{D}} \gamma, \quad \text{ii) } \beta_n \rightarrow \beta.$$

*Proof.* All that is left to do is show sufficiency. This is easy. Since

$$\int n \frac{y^2}{1+y^2} F_n(dy) = \gamma_n(+\infty) \leq M < \infty,$$

it follows that

$$\int \frac{y^2}{1+y^2} F_n(dy) \rightarrow 0;$$

hence  $F_n$  converges to the law degenerate at zero. Thus for all  $u$  in a finite interval, we can write

$$\log f_{S_n}(u) = (1 + \epsilon_n(u)) [i\beta_n u + \int \varphi(x, u)\gamma_n(dx)],$$

where  $\epsilon_n(u) \xrightarrow{uc} 0$ . Thus

$$\log f_{S_n}(u) \rightarrow i\beta\mu + \int \varphi(x, u)\gamma(dx).$$

Now we can go back and get the proof of 9.6. If  $S_n = X_1^{(n)} + \cdots + X_n^{(n)}$ , and  $S_n \xrightarrow{\mathcal{D}} X$ , then clearly  $X$  is normal if and only if  $\gamma_n$  converges to a measure  $\gamma$  concentrated on the origin. Equivalently, for every  $x > 0$ ,  $\gamma_n((-x, +x)^c) \rightarrow 0$ . Since

$$\gamma_n((-x, +x)^c) = \int_{|y| \geq x} \frac{y^2}{1+y^2} nF_n(dy),$$

this is equivalent to

$$nF_n((-x, +x)^c) \rightarrow 0, \quad x > 0 \quad \text{or} \quad nP(|X_1^{(n)}| \geq x) \rightarrow 0.$$

But

$$P(M_n < x) = P(|X_1^{(n)}| < x, \dots, |X_n^{(n)}| < x) = [1 - P(|X_1^{(n)}| \geq x)]^n.$$

Because  $X_1^{(n)} \xrightarrow{\mathcal{D}} 0$ ,

$$\log P(M_n < x) = -n(1 + \delta_n(x))P(|X_1^{(n)}| \geq x),$$

where  $\delta_n(x) \rightarrow 0$  for  $x > 0$ . Therefore,

$$M_n \xrightarrow{\mathcal{D}} 0 \Leftrightarrow nP(|X_1^{(n)}| \geq x) \rightarrow 0, \quad x > 0,$$

which completes the proof.

## 8. THE STABLE LAWS

Let  $X_1, X_2, \dots$  be identically distributed, nondegenerate, independent random variables. What is the class of all possible limit laws of normed sums

$$(9.23) \quad S_n = \frac{X_1 + \cdots + X_n}{A_n} - B_n, \quad A_n > 0?$$

Since  $S_n$  may be written as

$$S_n = X_1^{(n)} + \cdots + X_n^{(n)},$$

where

$$X_k^{(n)} = \frac{X_k}{A_n} - \frac{B_n}{n},$$

the requirement  $S_n \xrightarrow{\mathcal{D}} X$  implies that  $X$  is infinitely divisible. The condition  $X_1^{(n)} \xrightarrow{\mathcal{D}} 0$  implies  $A_n \rightarrow \infty$ ,  $B_n/n \rightarrow 0$ . This class of limit laws is the most interesting set of distributions following the normal and Poisson. Of course, if  $EX_1^2 < \infty$ , then  $A_n \sim \sqrt{n}$  and  $X$  must be normal. So the only interesting case is  $EX_1^2 = \infty$ . Two important questions arise.

*First:* What is the form of the class of all limit distributions  $X$  such that  $S_n \xrightarrow{\mathcal{D}} X$ ?

*Second:* Find necessary and sufficient conditions on the common distribution function of  $X_1, X_2, \dots$  so that  $S_n \xrightarrow{\mathcal{D}} X$ .

These two questions lead to the stable laws and the domains of attraction of the stable laws.

**Definition 9.24.** A random variable  $X$  is said to have a stable law if for every integer  $k > 0$ , and  $X_1, \dots, X_k$  independent with the same distribution as  $X$ , there are constants  $a_k > 0$ ,  $b_k$  such that

$$\mathcal{L}(X_1 + \cdots + X_k) = \mathcal{L}(a_k X + b_k).$$

This approach is similar to the way we intrinsically characterized the normal law; by breaking  $S_{nk}$  up into  $k$  blocks, we concluded that the limit of  $S_n/\sqrt{n}$  must satisfy  $\mathcal{L}(X_1 + \cdots + X_k) = \mathcal{L}(\sqrt{k}X)$ .

**Proposition 9.25.**  $X$  is the limit in distribution of normed sums (9.23) if and only if  $X$  has a stable law.

*Proof.* One way is quick: If  $X$  is stable, then  $\mathfrak{L}(X_1 + \cdots + X_n) = \mathfrak{L}(a_n X + b_n)$ . Then (check this by characteristic functions),

$$\mathfrak{L}(X) = \mathfrak{L}\left(\frac{X_1 + \cdots + X_n}{a_n} - \frac{b_n}{a_n}\right),$$

and we can take  $A_n = a_n$ ,  $B_n = b_n/a_n$  to get

$$\frac{X_1 + \cdots + X_n}{A_n} - B_n \xrightarrow{\mathfrak{D}} X$$

(actually  $\xrightarrow{\mathfrak{D}} X$  is true here).

To go the other way, suppose

$$Z_n = \frac{X_1 + \cdots + X_n}{A_n} - B_n \xrightarrow{\mathfrak{D}} X.$$

Then  $Z_{nk} \xrightarrow{\mathfrak{D}} X$  as  $n \rightarrow \infty$  for all  $k$ . Repeat the trick we used for the normal law:

$$Z_{nk} = \frac{1}{A_{nk}} (S_n^{(1)} + \cdots + S_n^{(k)}) - B_{nk},$$

where  $S_n^{(1)} = X_1 + \cdots + X_n$ ,  $S_n^{(2)} = X_{n+1} + \cdots + X_{2n}$ , ... Thus

$$\left(\frac{S_n^{(1)}}{A_n} - B_n\right) + \cdots + \left(\frac{S_n^{(k)}}{A_n} - B_n\right) = \frac{A_{nk}}{A_n} Z_{nk} + C_{n,k},$$

where  $C_{n,k} = (A_{nk}/A_n)B_{nk} - kB_n$ . By the law of convergence of types,

$$\frac{A_{nk}}{A_n} \rightarrow a_k, \quad C_{n,k} \rightarrow b_k.$$

Therefore,

$$\mathfrak{L}(X_1 + \cdots + X_k) = \mathfrak{L}(a_k X + b_k).$$

This not only proves 9.25 but contains the additional information that if  $S_n \xrightarrow{\mathfrak{D}} X$ , then  $A_{nk}/A_n \rightarrow a_k$  for all  $k$ . By considering the limit of  $A_{nmk}/A_n$  as  $n \rightarrow \infty$ , we conclude that the constants  $a_k$  must satisfy

$$(9.26) \quad a_{mk} = a_m a_k, \quad \text{all } k, m \geq 1.$$

## 9. THE FORM OF THE STABLE LAWS

**Theorem 9.27.** Let  $X$  have a stable law. Then either  $X$  has a normal distribution or there is a number  $\alpha$ ,  $0 < \alpha < 2$ , called the exponent of the law and constants  $m_1 \geq 0$ ,  $m_2 \geq 0$ ,  $\beta$  such that

$$\begin{aligned} \log f_X(u) = iu\beta + m_1 \int_0^\infty \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{dx}{x^{1+\alpha}} \\ + m_2 \int_{-\infty}^0 \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{dx}{|x|^{1+\alpha}}. \end{aligned}$$

*Proof.* Since  $X$  is stable, it is infinitely divisible,

$$\psi(u) = \log f_X(u) = iu\beta + \int \varphi(x, u)\gamma(dx).$$

In terms of characteristic function, the definition of stability becomes

$$[e^{\psi(u)}]^k = e^{ib_k u} e^{\psi(a_k u)},$$

or

$$(9.28) \quad k\psi(u) = ib_k u + \psi(a_k u).$$

Separate the situation into two cases.

$$\text{I. } \gamma(\{0\}) = 0, \quad \text{II. } \gamma(\{0\}) > 0.$$

CASE I. Define a measure  $\mu$ :

$$\mu(B) = \int_B \frac{1+x^2}{x^2} \gamma(dx).$$

Then  $\mu$  is  $\sigma$ -finite,  $\mu[-a, +a]^c < \infty$ , for any  $a > 0$ ,  $\int_{[-a, +a]} x^2 d\mu < \infty$ , and

$$\begin{aligned} \psi(u) &= iu\beta + \int \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \mu(dx), \\ \psi(a_k u) &= ia_k \beta u + \int \left( e^{iua_k x} - 1 - \frac{iua_k x}{1+x^2} \right) \mu(dx) \\ &= id_k u + \int \left( e^{iua_k x} - 1 - \frac{iua_k x}{1+a_k^2 x^2} \right) \mu(dx), \end{aligned}$$

where

$$d_k = a_k \beta + a_k \int \left[ \frac{x}{1+a_k^2 x^2} - \frac{x}{1+x^2} \right] \mu(dx).$$

This last integrand behaves like  $x^3$  near the origin, and is bounded away from the origin, so the integral exists. Define a change of variable measure  $\mu_k$  by

$$\mu_k(B) = \mu(z; a_k z \in B),$$

to get

$$\psi(a_k u) = id_k u + \int \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \mu_k(dx).$$

Therefore, (9.28) becomes

$$\begin{aligned} iu\beta k + \int \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) k\mu(dx) &= i(b_k + d_k)u \\ &\quad + \int \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \mu_k(dx). \end{aligned}$$

By the uniqueness of representation of infinitely divisible characteristic function we get the central result

$$(9.29) \quad k\mu \equiv \mu_k.$$

Let  $M^+(x) = \mu[x, \infty)$ ,  $x > 0$ . From (9.29),

$$(9.30) \quad kM^+(x) = M^+(x/a_k), \quad x > 0, k > 0.$$

Similarly, let  $M^-(x) = \mu(-\infty, x)$ ,  $x < 0$ . Again,  $kM^-(x) = M^-(x/a_k)$ ,  $x < 0$ ,  $k > 0$ .

**Proposition 9.31.**  $a_k = k^\lambda$ ,  $\lambda > 0$ , and

$$\begin{aligned} M^+(x) &= x^{-1/\lambda} M^+(1), \\ M^-(x) &= |x|^{-1/\lambda} M^-(-1). \end{aligned}$$

*Proof.*  $M^+(x)$  is nonincreasing. The relation  $kM^+(1) = M^+(1/a_k)$  implies  $a_k$  increasing in  $k$ , and we know  $a_k \rightarrow \infty$ . For any  $k$ ,  $a_n a_k = a_{nk}$  gives

$$a_{k^j} = (a_k)^j.$$

For  $n > k$ , take  $j$  such that  $k^j \leq n \leq k^{j+1}$ . Then

$$a_{k^j} \leq a_n \leq a_{k^{j+1}}$$

or

$$\begin{aligned} \log(a_{k^j}) &\leq \log a_n \leq \log(a_{k^{j+1}}), \\ j \log a_k &\leq \log a_n \leq (j+1) \log a_k. \end{aligned}$$

Dividing by  $j \log k$ , we get

$$\frac{\log a_k}{\log k} \leq \frac{\log a_n}{\log n} \left[ \frac{\log n}{j \log k} \right] \leq \frac{j+1}{j} \frac{\log a_k}{\log k}.$$

Now let  $n \rightarrow \infty$ ; consequently  $j \rightarrow \infty$ , and  $(\log n)/(j \log k) \rightarrow 1$ , implying

$$\frac{\log a_k}{\log k} = \lim_{n \rightarrow \infty} \frac{\log a_n}{\log n} = \lambda > 0.$$

To do the other part, set  $x = (k/n)^\lambda$ ; then in (9.30),

$$kM^+((k/n)^\lambda) = M^+(1/n^\lambda), \quad \text{all } k, n \geq 1.$$

For  $k = n$ , this is  $nM^+(1) = M^+(1/n^\lambda)$ . Substituting this above gives

$$kM^+((k/n)^\lambda) = nM^+(1)$$

or

$$M^+((k/n)^\lambda) = M^+(1)/(k/n).$$

For all  $x$  in the dense set  $\{(k/n)^\lambda\}$  we have shown  $M^+(x) = x^{-1/\lambda}M^+(1)$ . The fact that  $M^+(x)$  is nonincreasing makes this hold true for all  $x$ . Similarly for  $M^-(x)$ .

The condition  $\int_{-1}^1 x^2 d\mu < \infty$  implies  $\int_{-1}^1 x^2 |x|^{-(1/\lambda)-1} dx < \infty$  so that  $-1/\lambda + 1 > -1$ , or finally  $\lambda > \frac{1}{2}$ . For  $\psi(u)$  the expression becomes

$$\begin{aligned}\psi(u) = iu\beta + m_1 \int_0^\infty \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{dx}{x^{1+\alpha}} \\ + m_2 \int_{-\infty}^0 \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{dx}{|x|^{1+\alpha}},\end{aligned}$$

where  $m_1 = M^+(1) \cdot \frac{1}{\lambda}$ ,  $m_2 = M^-(-1) \cdot \frac{1}{\lambda}$ , and  $\alpha = \frac{1}{\lambda}$ ; so  $0 < \alpha < 2$ .

CASE II. If  $\gamma(\{0\}) = \sigma^2 > 0$ , then

$$\psi(u) = iu\beta - \frac{\sigma^2}{2} u^2 + \int_{\{0\}^c} \varphi(x, u) \gamma(dx).$$

The coefficient  $\sigma^2$  is uniquely determined by  $\lim_{u \rightarrow \infty} \psi(u)/u^2 = -\sigma^2/2$ , because  $\sup_{x, u} |\varphi(x, u)/u^2| < \infty$  and  $\varphi(x, u)/u^2 \rightarrow 0$  for  $x \neq 0$ , as  $u \rightarrow \infty$ . Apply the bounded convergence theorem to  $\int_{\{0\}^c} (\varphi(x, u)/u^2) \gamma(dx)$  to get the result. Therefore, dividing (9.28) by  $u^2$  and letting  $u \rightarrow \infty$  gives

$$k = a_k^2 \quad \text{which implies} \quad \lambda = \frac{1}{2}.$$

So (9.28) becomes

$$\begin{aligned}k\psi(u) &= ib_k u + \psi(\sqrt{k}u) \\ \psi(u) &= i \frac{b_k}{k} u + u^2 \frac{\psi(\sqrt{k}u)}{ku^2}.\end{aligned}$$

As  $k \rightarrow \infty$ ,  $\psi(\sqrt{k}u)/ku^2 \rightarrow -\sigma^2/2$ . This entails  $b_k/k \rightarrow \beta$  and

$$\psi(u) = iu\beta - (\sigma^2/2)u^2. \quad \text{Q.E.D.}$$

It is not difficult to check that every characteristic function of the form given in (9.27) is the characteristic function of a stable law. This additional fact completes our description of the form of the characteristic function for stable laws.

## Problems

2. Use the methods of this section to show that Proposition 9.1 holds without the restriction  $EX^2 < \infty$ .



3. Use Problem 2 to prove that if  $X_1, X_2, \dots$  are independent, identically distributed random variables and

$$\frac{X_1 + \dots + X_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} X,$$

then  $X$  is normal or degenerate.

4. Show that if  $\alpha < 1$ , then  $\psi(u)$  can be written as

$$\psi(u) = iu\beta + m_1 \int_0^\infty (e^{iux} - 1) \frac{dx}{x^{1+\alpha}} + m_2 \int_{-\infty}^0 (e^{iux} - 1) \frac{dx}{|x|^{1+\alpha}}.$$

Then prove that  $\beta = 0, m_2 = 0$ , implies  $X \geq 0$  a.s.

## 10. THE COMPUTATION OF THE STABLE CHARACTERISTIC FUNCTIONS

When the exponent  $\alpha$  is less than 2, the form of the stable characteristic function is given by 9.27. By doing some computations, we can evaluate these integrals in explicit form.

**Theorem 9.32.**  $f(u) = e^{\psi(u)}$  is the characteristic function of a stable law of exponent  $\alpha$ ,  $0 < \alpha < 1$ , and  $1 < \alpha < 2$  if and only if it has the form

$$\psi(u) = iuc - d|u|^\alpha \left( 1 + i\theta \frac{u}{|u|} \tan \frac{\pi}{2} \alpha \right),$$

where  $c$  is real,  $d$  real and positive, and  $\theta$  real such that  $|\theta| \leq 1$ . For  $\alpha = 1$ , the form of the characteristic function is given by

$$\psi(u) = iuc - d|u| \left( 1 + i\theta \frac{u}{|u|} \frac{2}{\pi} \log |u| \right),$$

with  $c, d, \theta$  as above.

*Proof.* Let

$$I_1(u) = \int_0^\infty \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{dx}{x^{1+\alpha}}.$$

Since  $I_1(-u) = \overline{I_1(u)}$ , we evaluate  $I_1(u)$  only for  $u > 0$ . Also,

$$I_2(u) = \int_{-\infty}^0 \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \frac{dx}{|x|^{1+\alpha}} = I_1(-u).$$

Consider first  $0 < \alpha < 1$ ; then

$$I_1(u) = \int_0^\infty \left( e^{iux} - 1 \right) \frac{dx}{x^{1+\alpha}} - iu \int_0^\infty \frac{1}{1+x^2} \frac{dx}{x^\alpha}.$$

Substitute  $ux = y$  in the first term to get, for  $u \geq 0$ ,

$$I_1(u) = iuc + u^\alpha H(\alpha),$$

where

$$H(\xi) = \int_0^\infty (e^{i y} - 1) \frac{dy}{y^{\xi+1}}.$$

For  $1 < \alpha < 2$ , integrate by parts, getting

$$\begin{aligned} I_1(u) &= \frac{i u}{\alpha} \int_0^\infty \left( e^{i u x} - \frac{d}{dx} \frac{x}{1+x^2} \right) \frac{dx}{x^\alpha} \\ &= iuc + \frac{i u}{\alpha} \int_0^\infty (e^{i u x} - 1) \frac{dx}{x^\alpha}, \end{aligned}$$

where

$$c = \frac{1}{\alpha} \int_0^\infty \left( 1 - \frac{d}{dx} \frac{x}{1+x^2} \right) \frac{dx}{x^\alpha}.$$

Substitute  $ux = y$  again, so

$$I_1(u) = iuc + u^\alpha \cdot \frac{i}{\alpha} H(\alpha - 1), \quad u \geq 0.$$

If  $\alpha = 1$ , the integration by parts gives

$$I_1(u) = i u \lim_{T \rightarrow \infty} \int_0^T \left( e^{i u x} - \frac{d}{dx} \frac{x}{1+x^2} \right) \frac{dx}{x}.$$

Let

$$J(T, u) = \int_0^T \left( e^{i u x} - \frac{d}{dx} \frac{x}{1+x^2} \right) \frac{dx}{x}.$$

Then for  $u_2 \geq u_1 > 0$ ,

$$\begin{aligned} J(T, u_2) - J(T, u_1) &= \int_0^T (e^{i u_2 x} - e^{i u_1 x}) \frac{dx}{x} \\ &= i \int_0^T dx \left[ \int_{u_1}^{u_2} e^{i v x} dv \right] = \int_{u_1}^{u_2} \left( \frac{e^{i v T} - 1}{v} \right) dv. \end{aligned}$$

Now, by the Riemann-Lebesgue lemma (see Chapter 10, Section 2),

$$\lim_T \int_{u_1}^{u_2} \frac{e^{i v T}}{v} dv = \lim_T \int e^{i v T} \chi_{[u_1, u_2]}(v) \cdot \frac{dv}{v} = 0$$

since  $\chi_{[u_1, u_2]}(v)/v$  is a bounded measurable function vanishing outside a finite interval. This gives

$$\lim_{T \rightarrow \infty} [J(T, u_2) - J(T, u_1)] = \log(u_1/u_2).$$

Consequently, checking that  $\lim_{T \rightarrow \infty} J(T, 1)$  exists,

$$\lim_{T \rightarrow \infty} J(T, u) = -\log u + c,$$

and

$$I_1(u) = iuc - iu \log u, \quad u \geq 0.$$

For the first time, the constant  $c$  appearing in the linear term is complex:

$$\begin{aligned} c &= \int_0^\infty \left( e^{ix} - \frac{d}{dx} \frac{x}{1+x^2} \right) \frac{dx}{x} \\ &= i \int_0^\infty \frac{\sin x}{x} dx + c_1, \end{aligned}$$

where  $c_1$  is real. The integral  $\int_0^\infty (\sin x/x) dx$  is well known and equals  $\pi/2$ . Finally, then,

$$I_1(u) = iuc_1 - (\pi/2)u - iu \log u, \quad u \geq 0.$$

The remaining piece of work is to evaluate

$$H(\xi) = \int_0^\infty (e^{i\xi y} - 1) \frac{dy}{y^{\xi+1}}, \quad 0 < \xi < 1.$$

This can be done by contour integration (see Gnedenko and Kolmogorov, [62, p. 169]), with the result that

$$H(\xi) = e^{-i(\pi/2)\xi} L(\xi),$$

where  $L(\xi)$  is real and negative. Putting everything together, we get

$$m_1 I_1(u) + m_2 I_1(-u) = m_1 I_1(u) + m_2 \overline{I_1(u)}.$$

For  $0 < \alpha < 1$ , and  $n \geq 0$ ,

$$\begin{aligned} \psi(u) &= icu + u^\alpha (m_1 H(\alpha) + m_2 \overline{H(\alpha)}) \\ &= icu - du^\alpha (1 + i\theta \tan(\pi/2)\alpha), \end{aligned}$$

where  $d = -(m_1 + m_2)R/H(\alpha)$  is real and positive,  $\theta = (m_1 - m_2)/(m_1 + m_2)$  is real with range  $[-1, +1]$ , and  $c$  is real. For  $1 < \alpha < 2$ ,

$$\psi(u) = icu + \frac{iu^\alpha}{\alpha} (m_1 H(\alpha - 1) - m_2 \overline{H(\alpha - 1)}).$$

Now

$$H(\alpha - 1) = e^{-i(\pi/2)(\alpha-1)} L(\alpha - 1) = i e^{-i(\pi/2)\alpha} L(\alpha - 1),$$

so

$$\psi(u) = icu - du^\alpha (1 + i\theta \tan(\pi/2)\alpha), \quad u \geq 0.$$

Here  $d = ((m_1 + m_2)/\alpha) \operatorname{Re}(e^{-i(\pi/2)\alpha} L(\alpha - 1))$  is real and positive, and  $\theta$  is again  $(m_1 - m_2)/(m_1 + m_2)$  with range  $[-1, +1]$ . If  $\alpha = 1$ , then

$$\begin{aligned}\psi(u) &= icu - (\pi/2)u(m_1 + m_2) + iu(m_1 - m_2) \log u \\ &= icu - du(1 + i\theta \cdot (2/\pi) \log u), \quad u \geq 0,\end{aligned}$$

$\theta$  as above,  $d$  real and positive,  $c$  real. Q.E.D.

**Problem 5.** Let  $S_n = X_1 + \cdots + X_n$  be consecutive sums of independent random variables each having the same symmetric stable law of exponent  $\alpha$ . Use the Fourier inversion theorem and the technique of Problem 24, Chapter 8, to show that the sums are transient for  $0 < \alpha < 1$ , recurrent for  $\alpha \geq 1$ . The case  $\alpha = 1$  provides an example where  $E|X_1| = \infty$ , but the sums are recurrent. Show that for all  $\alpha > 0$ ,

$$P(\overline{\lim} S_n = \infty) = P(\overline{\lim} S_n = -\infty) = 1.$$

Conclude that the sums change sign infinitely often for any  $\alpha > 0$ .

## 11. THE DOMAIN OF ATTRACTION OF A STABLE LAW

Let  $X_1, X_2, \dots$  be independent and identically distributed. What are necessary and sufficient conditions on their distribution function  $F(x)$  such that  $S_n$  suitably normalized converges in distribution to  $X$  where  $X$  is nondegenerate? Of course, the limit random variable  $X$  must be stable.

**Definition 9.33.** The distribution  $F(x)$  is said to be in the domain of attraction of a stable law with exponent  $\alpha < 2$  if there are constants  $A_n, B_n$  such that

$$\frac{S_n - B_n}{A_n} \xrightarrow{\mathcal{D}} X$$

and  $X$  has exponent  $\alpha$ . Denote this by  $F \in D(\alpha)$ .

Complete conditions on  $F(x)$  are given by

**Theorem 9.34.**  $F(x)$  is in the domain of attraction of a stable law with exponent  $\alpha < 2$  if and only if there are constants  $M^+, M^- \geq 0$ ,  $M^+ + M^- > 0$ , such that as  $y \rightarrow \infty$ :

$$\text{i) } \lim_{y \rightarrow \infty} \frac{F(-y)}{1 - F(y)} = \frac{M^-}{M^+};$$

ii) For every  $\xi > 0$ ,

$$M^+ > 0 \Rightarrow \lim_{y \rightarrow \infty} \frac{1 - F(\xi y)}{1 - F(y)} = \frac{1}{\xi^\alpha},$$

$$M^- > 0 \Rightarrow \lim_{y \rightarrow \infty} \frac{F(-\xi y)}{F(-y)} = \frac{1}{\xi^\alpha}.$$

*Proof.* We show first *necessity*. By Theorem 9.22,  $\gamma_n \xrightarrow{\mathcal{D}} \gamma$ . Now take  $b_n = B_n/n$ . Then

$$\gamma_n(B) = n \int_B \frac{y^2}{1 + y^2} F_n(dy),$$

where

$$F_n(x) = P\left(\frac{X_1}{A_n} - b_n < x\right) = F(A_n(x + b_n)),$$

and

$$\gamma(B) = \int_B \frac{y^2}{1 + y^2} \mu(dy).$$

Thus, for any  $x_1 < 0$ ,  $x_2 > 0$ , since the  $\gamma_n$  must be mass-preserving,

$$\int_{[x_1, x_2]^c} \frac{1 + z^2}{z^2} \gamma_n(dz) \rightarrow \int_{[x_1, x_2]^c} \frac{1 + z^2}{z^2} \gamma(dz).$$

This is

$$n \int_{[x_1, x_2]^c} F_n(dz) \rightarrow \int_{[x_1, x_2]^c} \mu(dz).$$

By taking one or the other of  $x_1$ ,  $x_2$  to be infinite, we find that the condition becomes

$$\begin{aligned} n[1 - F_n(x)] &\rightarrow M^+(x), \quad x > 0, \\ nF_n(x) &\rightarrow M^-(x), \quad x < 0. \end{aligned}$$

Then

$$\begin{aligned} n[1 - F_n(A_n(x + b_n))] &\rightarrow M^+(x), \quad x > 0, \\ nF_n(A_n(x + b_n)) &\rightarrow M^-(x), \quad x < 0. \end{aligned}$$

Since  $b_n \rightarrow 0$ , for any  $\epsilon > 0$  and  $n$  sufficiently large,

$$F(A_n(x - \epsilon + b_n)) \leq F(A_n x) \leq F(A_n(x + \epsilon + b_n)).$$

We can use this to get

$$\begin{aligned} M^+(x + \epsilon) &\leq \varlimsup n[1 - F(A_n x)] \\ &\leq \varlimsup n[1 - F(A_n x)] \leq M^+(x - \epsilon), \quad x > 0, \\ M^-(x - \epsilon) &\leq \varlimsup nF(A_n x) \\ &\leq \varlimsup nF(A_n x) \leq M^-(x + \epsilon), \quad x < 0. \end{aligned}$$

We know that  $M^+(x)$ ,  $M^-(x)$  are continuous, so we can conclude that

$$\begin{aligned} n[1 - F(A_n x)] &\rightarrow M^+(x), \quad x > 0, \\ nF(A_n x) &\rightarrow M^-(x), \quad x < 0. \end{aligned}$$

Now if we fix  $x > 0$ , for any  $y > 0$  sufficiently large, there is an  $n$  such that

$$A_n x \leq y \leq A_{n+1} x.$$

Then

$$F(A_n x) \leq F(y) \leq F(A_{n+1} x),$$

$$F(-A_{n+1} x) \leq F(-y) \leq F(-A_n x).$$

So for any  $\xi > 0$ ,

$$\frac{1 - F(A_{n+1} \xi x)}{1 - F(A_n x)} \leq \frac{1 - F(\xi y)}{1 - F(y)} \leq \frac{1 - F(A_n \xi x)}{1 - F(A_{n+1} x)},$$

$$\frac{M^+(\xi x)}{M^+(x)} \leq \lim_{y \rightarrow \infty} \frac{1 - F(\xi y)}{1 - F(y)} \leq \overline{\lim}_{y \rightarrow \infty} \frac{1 - F(\xi y)}{1 - F(y)} \leq \frac{M^+(\xi x)}{M^+(x)}$$

$$\Rightarrow \lim_{y \rightarrow \infty} \frac{1 - F(\xi y)}{1 - F(y)} = \frac{1}{\xi^\alpha}.$$

In exactly the same way, conclude

$$\lim_{y \rightarrow \infty} \frac{F(-\xi y)}{F(-y)} = \frac{1}{\xi^\alpha}.$$

Also,

$$\frac{F(-A_{n+1} x)}{1 - F(A_n x)} \leq \frac{F(-y)}{1 - F(y)} \leq \frac{F(-A_n x)}{1 - F(A_{n+1} x)},$$

which leads to

$$\lim_{x \rightarrow \infty} \frac{F(-y)}{1 - F(y)} = \frac{M^-(-x)}{M^+(x)} = \frac{M^-(-1)}{M^+(1)}.$$

To get *sufficiency*, assume, for example, that  $M^+ > 0$ . I assert we can define constants  $A_n > 0$  such that

$$n(1 - F(A_n)) \rightarrow M^+.$$

Condition (ii) implies  $1 - F(x) > 0$ , all  $x > 0$ . Take  $A_n$  such that for any  $\epsilon > 0$ ,  $n(1 - F(A_n)) \geq M^+$ , but  $n(1 - F(A_n + \epsilon)) \leq M^+$ . Then if  $\overline{\lim} n(1 - F(A_n)) = M^+(1 + \delta)$ ,  $\delta > 0$ , there is a subsequence  $n'$  such that for every  $\epsilon > 0$ ,

$$\overline{\lim} \frac{1 - F(A_{n'} + \epsilon)}{1 - F(A_{n'})} \leq \frac{1}{1 + \delta}.$$

This is ruled out by condition (ii). So

$$n(1 - F(A_n x)) = n(1 - F(A_n)) \cdot \frac{1 - F(A_n x)}{1 - F(A_n)} \rightarrow M^+ \cdot \frac{1}{x^\alpha}.$$

Similarly,

$$nF(-A_n x) \rightarrow M^- \cdot \frac{1}{x^\alpha}.$$

Take  $F$  to be the distribution function of a random variable  $X_1$ , and write

$$\mu_n(B) = nP\left(\frac{X_1}{A_n} \in B\right), \quad B \in \mathfrak{B}_1.$$

Then

$$\begin{aligned}\mu_n([x, \infty)) &\rightarrow \frac{M^+}{x^\alpha}, \quad x > 0, \\ \mu_n((-\infty, x)) &\rightarrow \frac{M^-}{|x|^\alpha}, \quad x < 0.\end{aligned}$$

Therefore, the  $\{\mu_n\}$  sequence is mass-preserving, and so is the  $\{\gamma_n\}$  sequence defined, as before, by

$$\gamma_n(B) = \int_B \frac{y^2}{1+y^2} \mu_n(dy), \quad B \in \mathfrak{B}_1.$$

Take  $f(x)$  any bounded continuous function, and  $\epsilon > 0$ :

$$\int f(x) \gamma_n(dx) = \int_{(-\epsilon, +\epsilon)^c} f(x) \gamma_n(dx) + \int_{(-\epsilon, +\epsilon)} f(x) \gamma_n(dx).$$

Let  $g(x) = f(x)x^2/(1+x^2)$ . The first integral converges to

$$\begin{aligned}\alpha M^+ \int_\epsilon^\infty g(x) \frac{dx}{x^{\alpha+1}} + \alpha M^- \int_{-\infty}^{-\epsilon} g(x) \frac{dx}{|x|^{\alpha+1}} \\ = \alpha M^+ \int_0^\infty g(x) \frac{dx}{x^{\alpha+1}} + \alpha M^- \int_{-\infty}^0 g(x) \frac{dx}{|x|^{\alpha+1}} + \delta_1(\epsilon),\end{aligned}$$

where  $\delta_1(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ . Thus, defining  $\gamma(x)$  in the obvious way,

$$\overline{\lim} \left| \int f(x) \gamma_n(dx) - \int f(x) \gamma(dx) \right| \leq \delta_1(\epsilon) + \overline{\lim} \int_{|x| < \epsilon} |f(x)| \gamma_n(dx).$$

To complete the proof that  $\gamma_n \xrightarrow{\mathfrak{D}} \gamma$ , we need

**Proposition 9.35**

$$\overline{\lim} \int_{|x| < \epsilon} x^2 \mu_n(dx) = \delta_2(\epsilon),$$

where  $\delta_2(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ .

In order not to interrupt the main flow of the proof, I defer the proof of 9.35 until the end. Since  $A_n \rightarrow \infty$ , the characteristic function  $g_n(u)$  of  $X_1 + \cdots + X_n/A_n$  is given, as before, by

$$\log g_n(u) = (1 + \epsilon_n(u)) \int \varphi(x, u) \gamma_n(dx) + (1 + \epsilon_n(u)) i u \beta_n,$$

where

$$\beta_n = \int \frac{x}{1+x^2} \mu_n(dx)$$

and  $\epsilon_n(u) \xrightarrow{uc} 0$ . Since  $\gamma_n \xrightarrow{\mathcal{D}} \gamma$ , the first term tends toward  $\int \varphi(x, u) \gamma(dx)$ . The characteristic function of  $S_n/A_n - \beta_n$  is  $e^{-iu\beta_n} g_n(u)$ . So, if  $\epsilon_n(u)\beta_n \rightarrow 0$ , then the theorem (except for 9.35) follows with  $B_n = \beta_n$ . For  $n$  sufficiently large,

$$|\epsilon_n(u)| \leq |1 - f(u/A_n)|,$$

where  $f$  is the characteristic function of  $X_1$ . But

$$f\left(\frac{u}{A_n}\right) - 1 = \frac{1}{n} \int (e^{iu x} - 1) \mu_n(dx) = \frac{1}{n} \left[ \int \varphi(x, u) \gamma_n(dx) + iu\beta_n \right],$$

so it is sufficient to show that  $\beta_n^2/n \rightarrow 0$ . For any  $\epsilon > 0$ , use the Schwarz inequality to get

$$\begin{aligned} \overline{\lim} \frac{\beta_n^2}{n} &\leq \overline{\lim} \frac{1}{n} \left[ \int_{|x| < \epsilon} |x| \mu_n(dx) \right]^2 \\ &\leq \overline{\lim} \left[ \frac{1}{n} \mu_n((-\epsilon, +\epsilon)) \int_{|x| < \epsilon} x^2 \mu_n(dx) \right] \\ &\leq \overline{\lim} \int_{|x| < \epsilon} x^2 \mu_n(dx). \end{aligned}$$

Apply 9.35 to reach the conclusion.

*Proof of 9.35.* We adapt a proof due to Feller [59, Vol. II]. Write

$$I(x) = \int_0^x y(1 - F(y)) dy.$$

We begin by showing that there is a constant  $c$  such that

$$I(x) \leq cx^2(1 - F(x)).$$

For  $t > \tau$ ,

$$(9.36) \quad I(tx) = I(\tau x) + x^2 \int_{\tau}^t \xi(1 - F(x\xi)) d\xi.$$

Fix  $x > 1$ . For any  $\epsilon > 0$ , take  $\tau$  so large that for  $\xi > \tau$ ,

$$\frac{1 - F(x\xi)}{1 - F(\xi)} \geq (1 - \epsilon)x^{-\alpha}.$$

From (9.36)

$$I(tx) \geq I(\tau x) + x^{2-\alpha}(1 - \epsilon)(I(t) - I(\tau)).$$

This inequality implies  $I(t) \rightarrow \infty$  as  $t \rightarrow \infty$ . Then, dividing by  $I(t)$  and letting  $t \rightarrow \infty$  yields

$$\underline{\lim} \frac{I(tx)}{I(t)} \geq x^{2-\alpha}.$$



Now in 9.36 take  $\tau = 1$ , so

$$\frac{I(tx)}{I(x)} = 1 + \frac{x^2(1 - F(x))}{I(x)} \int_1^t \frac{1 - F(x\xi)}{1 - F(x)} \cdot \xi d\xi.$$

The integrand is bounded by  $\xi$ , so

$$\frac{I(tx)}{I(x)} \leq 1 + t^2 \frac{x^2(1 - F(x))}{I(x)}, \quad t > 1.$$

Thus, taking  $\liminf$  on  $x$ ,

$$t^{2-\alpha} \leq 1 + t^2 \liminf \frac{x^2(1 - F(x))}{I(x)}.$$

This does it.

To relate this to 9.35 integrate by parts on  $\int_{[0, \epsilon]} x^2 \mu_n(dx)$ , assuming  $\epsilon$  is a continuity point of  $\mu_n$  for all  $n$ . This gives

$$\int_{[0, \epsilon]} x^2 \mu_n(dx) \leq 2 \int_{[0, \epsilon]} x \mu_n([x, \infty)) dx.$$

From  $\mu_n([x, \infty)) = n(1 - F(A_n x))$ , we get

$$(9.37) \quad \int_{[0, \epsilon]} x^2 \mu_n(dx) \leq \frac{2n}{A_n^2} \int_0^{\epsilon A_n} y(1 - F(y)) dy = \frac{2n}{A_n^2} I(\epsilon A_n).$$

Because of the inequality proved above,

$$\int_{[0, \epsilon]} x^2 \mu_n(dx) \leq 2nc\epsilon^2(1 - F(A_n\epsilon)).$$

Therefore,

$$\liminf \int_{[0, \epsilon]} x^2 \mu_n(dx) \leq 2cM^+ \epsilon^{2-\alpha}.$$

The integral over the range  $(-\epsilon, 0)$  is treated in the same way, with  $F(-x)$  in place of  $1 - F(x)$ . Q.E.D.

## Problems

6. Show that  $n/A_n^2 \rightarrow 0$ . [For any  $\delta > 0$ , take  $x_0$  such that for  $x \geq x_0$ ,

$$\frac{1 - F(2x)}{1 - F(x)} \geq 2^{-(\alpha+\delta)}.$$

Define  $k$  by  $2^{k-1} < A_n \leq 2^k$ ; then

$$1 - F(A_n) \geq 1 - F(2^k) = \prod_{j=m}^k \frac{1 - F(2^j)}{1 - F(2^{j-1})} (1 - F(2^{m-1})).$$

Now select  $m$  appropriately.]

7. Show, using the same methods used to get  $n/A_n^2 \rightarrow 0$ , that for  $F(x) \in D(\alpha)$ ,

$$\lim_{x \rightarrow \infty} x^\lambda [1 - F(x)] = \begin{cases} \infty, & \lambda > \alpha, \\ 0, & \lambda < \alpha. \end{cases}$$

Conclude that

$$\int |x|^\lambda F(dx) \begin{cases} < \infty, & \lambda < \alpha, \\ = \infty, & \lambda > \alpha. \end{cases}$$

8. For  $F(x) \in D(\alpha)$ ,  $\alpha < 1$ , by using the same methods as in the proof of 9.35, show that

$$\overline{\lim} \int_{|x| \leq \epsilon} |x| \mu_n(dx) = \delta(\epsilon), \quad \delta(\epsilon) \rightarrow 0 \quad \text{as } \epsilon \rightarrow 0.$$

Conclude that  $\beta_n$  converges to a finite constant  $\beta$ ; hence  $B_n$  can be defined to be zero, all  $n$ .

9. For  $F(x) \in D(\alpha)$ ,  $1 < \alpha < 2$ , show that

$$\int \frac{x}{1+x^2} \mu_n(dx) - \int x \mu_n(dx)$$

converges, so  $B_n/n$  can be taken as  $-EX_1/A_n$ ; that is,  $X_1, X_2, \dots$  are replaced by  $X_1 - EX_1, X_2 - EX_2, \dots$

10. For  $F(x) \in D(\alpha)$ ,  $\alpha < 1$ , if  $F(x) = 0$ ,  $x \leq 0$ , then

$$\log f_{S_n/A_n}(u) = \int_0^\infty (e^{iux} - 1) \mu_n(dx) \cdot (1 + \epsilon_n(u)).$$

Prove, using Problem 8 and the computations of the various integrals defining the stable laws, that

$$\frac{S_n}{A_n} \xrightarrow{\mathcal{D}} X, \quad \text{where} \quad \log f_X(u) = M^+ L(\alpha) e^{-i(\pi/2)\alpha} u^\alpha, \quad u \geq 0.$$

## 12. A COIN-TOSSING EXAMPLE

The way to recognize laws in  $D(\alpha)$  is

**Definition 9.38.** A function  $H(x)$  on  $(0, \infty)$  is said to be slowly changing if for all  $\xi \in (0, \infty)$ ,

$$\lim_{x \rightarrow \infty} \frac{H(\xi x)}{H(x)} = 1.$$

For example,  $\log x$  is slowly changing, so is  $\log \log x$ , but not  $x^\alpha$ ,  $\alpha \neq 0$ .

**Proposition 9.39.**  $F(x) \in D(\alpha)$  if and only if, defining  $H^+(x)$ ,  $H^-(x)$  on  $(0, \infty)$  by

$$1 - F(x) = \frac{1}{x^\alpha} H^+(x), \quad F(-x) = \frac{1}{x^\alpha} H^-(x),$$

there are constants  $M^+ \geq 0$ ,  $M^- \geq 0$ ,  $M^+ + M^- > 0$  such that

$M^+ > 0 \Rightarrow H^+(x)$  slowly changing,  $M^- > 0 \Rightarrow H^-(x)$  slowly changing, and, as  $x \rightarrow \infty$ ,

$$\frac{H^+(x)}{H^-(x)} \rightarrow \frac{M^+}{M^-}.$$

*Proof.* Obvious.

Now, in fair coin-tossing, if  $T_1$  is the time until the first return to equilibrium, then by Problem 19, Chapter 3,

$$P(T_1 > n) = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{n^{1/2}} (1 + \delta_n),$$

where  $\delta_n \rightarrow 0$ , so  $1 - F(n) \sim cn^{-1/2}$ . Thus for the range  $n \leq x \leq n+1$ ,  $1 - F(x) \sim cn^{-1/2}$  or finally,  $1 - F(x) = cx^{-1/2}(1 + \delta(x))$ ,  $\delta(x) \rightarrow 0$ . Therefore, by 9.39,  $F(x) \in D(\frac{1}{2})$ . To get the normalizing  $A_n$ , the condition is  $n(1 - F(A_n))$  converges to  $M^+$ , that is,

$$cn \cdot \frac{1}{\sqrt{A_n}} (1 + \delta(A_n)) \rightarrow M^+.$$

Take  $A_n = n^2$ ; then  $M^+ = c = \sqrt{2}/\sqrt{\pi}$ . We conclude, if  $R_n$  is the time of the  $n$ th return to equilibrium, then

**Theorem 9.40.**  $R_n/n^2 \xrightarrow{D} X$ , where

$$\log f_X(u) = \frac{\sqrt{2}}{\sqrt{\pi}} u^\alpha e^{-i(\pi/2)\alpha} L(\alpha), \quad \alpha = \frac{1}{2}, \quad u \geq 0.$$

*Proof.* See Problem 10.

### 13. THE DOMAIN OF ATTRACTION OF THE NORMAL LAW

If we look at the various possible distribution functions  $F(x)$  and ask when can we normalize  $S_n$  so that  $S_n/A_n - B_n$  converges to a nondegenerate limit, we now have some pretty good answers. For  $F(x)$  to be in  $D(\alpha)$ , the tails of the distribution must behave in a very smooth way—actually the mass of the distribution out toward  $\infty$  must mimic the behavior of the tails of the limiting stable law. So for  $0 < \alpha < 2$ , only a few distribution functions  $F(x)$  are in  $D(\alpha)$ . But the normal law is the limit for a wide class of distributions, including all  $F(x)$  such that  $\int x^2 F(dx) < \infty$ . The obvious unanswered question is: What else does this class contain? In other words, for what distributions  $F(x)$  does  $S_n/A_n - B_n \xrightarrow{D} \mathcal{N}(0, 1)$  for an appropriate choice of  $A_n, B_n$ ?

We state the result only—see Gnedenko and Kolmogorov, [62, pp. 172 ff.], for the proof.

**Theorem 9.41.** *There exist  $A_n, B_n$  such that  $S_n/A_n - B_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$  if and only if*

$$(9.42) \quad \lim_{x \rightarrow \infty} \frac{x^2 \int_{|y| > x} F(dy)}{\int_{|y| \leq x} y^2 F(dy)} = 0.$$

**Problem 11.** Show that  $\int y^2 F(dy) < \infty$  implies that the limit in (9.42) is zero. Find a distribution such that  $\int y^2 F(dy) = \infty$ , but (9.42) is satisfied.

## NOTES

The central limit theorem 9.2 dates back to Lyapunov [109, 1901]. The general setting of the problem into the context of infinitely divisible laws starts with Kolmogorov [94, 1932], who found all infinitely divisible laws with finite second moment, and Paul Lévy [102, 1934], who derived the general expression while investigating processes depending on a continuous-time parameter (see Chapter 14). The present framework dates to Feller [51] and Khintchine [89] in 1937. Stable distributions go back to Paul Lévy [100, 1924], and also [103, 1937]. In 1939 and 1940 Doeblin, [27] and [29], analyzed the problem of domains of attraction. One fascinating discovery in his later paper was the existence of universal laws. These are laws  $\mathcal{L}(X)$  such that for  $S_n = X_1 + \cdots + X_n$  sums of independent random variables each having the law  $\mathcal{L}(X)$ , there are normalizing constants  $A_n, B_n$  such that for  $Y$  having any infinitely divisible distribution, there is a subsequence  $\{n_k\}$  with

$$\frac{S_{n_k} - B_{n_k}}{A_{n_k}} \xrightarrow{\mathcal{D}} Y.$$

For more discussion of stable laws see Feller's book [59, Vol. II]. A much deeper investigation into the area of this chapter is given by Gnedenko and Kolmogorov [62].

## CHAPTER 10

# THE RENEWAL THEOREM AND LOCAL LIMIT THEOREM

### 1. INTRODUCTION

By sharpening our analytic tools, we can prove two more important weak limit theorems regarding the distribution of sums of independent, identically distributed random variables. We group these together because the methods are very similar, involving a more delicate use of characteristic functions and Fourier analytical methods. In the last sections, we apply the local limit theorem to get occupation time laws. This we do partly because of their own interest, and also because they illustrate the use of Tauberian arguments and the method of moments.

### 2. THE TOOLS

A basic necessity is the

**Riemann-Lebesgue lemma.** Let  $f(x)$  be  $\mathcal{B}_1$ -measurable and  $\int |f(x)| dx < \infty$ ; then

$$(10.1) \quad \lim_{u \rightarrow \pm\infty} \int e^{iux} f(x) dx = 0.$$

*Proof.* For any  $\epsilon > 0$ , take  $I$  a finite interval such that

$$\int_I |f(x)| \leq \epsilon$$

and take  $M$  such that

$$\int_{\{|f(x)| > M\}} |f(x)| dx \leq \epsilon.$$

Therefore,

$$\left| \int e^{iux} f(x) dx - \int_{I \cap \{|f(x)| \leq M\}} e^{iux} f(x) dx \right| \leq 2\epsilon,$$

so it is sufficient to prove this lemma for bounded  $f(x)$  vanishing off of finite intervals  $I$ . Then (see Problem 5, Chapter 5) for any  $\epsilon > 0$ , there are disjoint intervals  $I_1, \dots, I_n$  such that

$$\int \left| f(x) - \sum_k \alpha_k \chi_{I_k}(x) \right| dx \leq \epsilon.$$

By direct computation,

$$\int e^{iux} \chi_{I_k}(x) dx \rightarrow 0$$

as  $u$  goes to  $\pm \infty$ .

Next, we need to broaden the concept of convergence in distribution. Consider the class  $\mathcal{Q}$  of  $\sigma$ -additive measures  $\mu$  on  $\mathcal{B}_1$  such that  $\mu(I)$  is finite for every finite interval  $I$ .

**Definition 10.2.** Let  $\mu_n, \mu \in \mathcal{Q}$ . Say that  $\mu_n$  converges weakly to  $\mu$ ,  $\mu_n \xrightarrow{w} \mu$ , if for every continuous function  $\varphi(x)$  vanishing off of a finite interval

$$(10.3) \quad \int \varphi(x) \mu_n(dx) \rightarrow \int \varphi(x) \mu(dx).$$

If the  $\mu_n, \mu$  have total mass one, then weak convergence coincides with convergence in distribution. Some of the basic results concerning convergence in distribution have easy extensions.

**Proposition 10.4.**  $\mu_n \xrightarrow{w} \mu$  iff for every Borel set  $A$  contained in a finite interval such that  $\mu(\text{bd}(A)) = 0$ ,

$$\mu_n(A) \rightarrow \mu(A).$$

The proof is exactly the same as in Chapter 8. The Helly-Bray theorem extends to

**Theorem 10.5.** If  $\mu_n \in \mathcal{Q}$  and for every finite interval  $I$ ,  $\overline{\lim} \mu_n(I) < \infty$ , then there is a subsequence  $\mu_{n_k}$  converging weakly to  $\mu \in \mathcal{Q}$ .

*Proof.* For  $I_k = [-k, +k]$ ,  $k = 1, 2, \dots$ , use the Helly-Bray theorem to get an ordered subset  $N_k$  of the positive integers such that on the interval  $I_k$ ,  $\mu_n \xrightarrow{w} \mu^{(k)}$  as  $n$  runs through  $N_k$ , and  $N_{k+1} \subset N_k$ . Here we use the obvious fact that the Helly-Bray theorem holds for measures whose total mass is bounded by the same constant. Let  $\mu$  be the measure that agrees with  $\mu^{(k)}$  on Borel subsets of  $I_k$ . Since  $N_{k+1} \subset N_k$ ,  $\mu$  is well defined. Let  $n_k$  be the  $k$ th member of  $N_k$ . Then clearly,  $\mu_{n_k} \xrightarrow{w} \mu$ .

There are also obvious generalizations of the ideas of separating classes of functions. But the key result we need is this:

**Definition 10.6.** Let  $\mathcal{H}$  be the class of  $\mathcal{B}_1$ -measurable complex-valued functions  $h(x)$  such that

$$\int |h(x)| dx < \infty$$

and

$$h(x) = \int e^{iux} \hat{h}(u) du,$$

where  $\hat{h}(u)$  is real and vanishes outside of a finite interval  $I$ .

Note that if  $h \in \mathcal{H}$ , then for any real  $v$ , the function  $e^{ivx}h(x)$  is again in  $\mathcal{H}$ .

**Theorem 10.7.** Let  $\mu_n, \mu \in \mathfrak{Q}$ . Suppose that there is an everywhere positive function  $h_0 \in \mathcal{K}$  such that  $\int h_0 d\mu$  is finite, and

$$\int h d\mu_n \rightarrow \int h d\mu$$

for all functions  $h \in \mathcal{K}$  of the form  $e^{ivx} h_0(x)$ ,  $v$  real. Then  $\mu_n \xrightarrow{w} \mu$ .

*Proof.* Let  $\alpha_n = \int h_0 d\mu_n$ ,  $\alpha = \int h_0 d\mu$ . Note that  $h_0$ , or for that matter any  $h \in \mathcal{K}$ , is continuous. If  $\alpha = 0$ , then  $\mu \equiv 0$ , and since  $\alpha_n \rightarrow \alpha$ , for any finite interval  $I$ ,  $\mu_n(I) \rightarrow 0$ . Hence, assume  $\alpha > 0$ . Define probability measures  $\nu_n, \nu$  on  $\mathcal{B}_1$  by

$$\nu_n(B) = \frac{1}{\alpha_n} \int_B h_0(x) \mu_n(dx), \quad \nu(B) = \frac{1}{\alpha} \int_B h_0(x) \mu(dx).$$

By the hypothesis of the theorem, for all real  $v$ ,

$$\int e^{ivx} \nu_n(dx) \rightarrow \int e^{ivx} \nu(dx),$$

so that  $\nu_n \xrightarrow{w} \nu$ . Thus, by 8.12, for any bounded continuous  $g(x)$ ,

$$\int g(x) h_0(x) \mu_n(dx) \rightarrow \int g(x) h_0(x) \mu(dx).$$

For any continuous  $\varphi(x)$  vanishing off of a finite interval  $I$ , take  $g(x) = \varphi(x)/h_0(x)$  to conclude

$$\int \varphi(x) \mu_n(dx) \rightarrow \int \varphi(x) \mu(dx).$$

A question remains as to whether  $\mathcal{K}$  contains any everywhere positive functions. To see that it does, check that

$$h_\lambda(x) = \left( \frac{\sin \lambda x}{\lambda x} \right)^2, \quad \lambda > 0$$

in  $\mathcal{K}$ , with

$$\hat{h}_\lambda(u) = \begin{cases} (2\lambda - |u|)/4\lambda^2, & |u| \leq 2\lambda, \\ 0, & |u| > 2\lambda. \end{cases}$$

Now, for  $\lambda_1, \lambda_2$  not rational multiples of each other, the function  $h_0(x) = h_{\lambda_1}(x) + h_{\lambda_2}(x)$  is everywhere positive and in  $\mathcal{K}$ .

### 3. THE RENEWAL THEOREM

For independent, identically distributed random variables  $X_1, X_2, \dots$ , define, as usual,  $S_n = X_1 + \dots + X_n$ ,  $S_0 = 0$ . In the case that the sums are transient, so that  $P(S_n \in I \text{ i.o.}) = 0$ , all finite  $I$ , there is one major result

concerning the interval distribution of the sums. It interests us chiefly in the case of finite nonzero first moment, say  $0 < EX_1 < \infty$ . The law of large numbers guarantees  $S_n \rightarrow \infty$  a.s. But there is still the possibility of a more or less regular progression of the sums out to  $\infty$ . Think of  $X_1, X_2, \dots$  as the successive life spans of light bulbs in a given socket. After a considerable number of years of operation, the distributions of this replacement process should be invariant under shifts of time axis. For instance, the expected number of failures in any interval of time  $I$  should depend only on the length of  $I$ . This is essentially the renewal theorem: Let  $l_d$  assign mass  $d$  to every point of  $L_d$ ,  $l_0$  is Lebesgue measure on  $L_0$ .

**Theorem 10.8.** Suppose  $X_1$  is distributed on the lattice  $L_d$ ,  $d \geq 0$ . Let  $N(I)$  be the number of members of the sequence  $S_0, S_1, S_2, \dots$  landing in the finite interval  $I$ . Then as  $y \rightarrow \infty$  through  $L_d$ ,

$$\lim_y EN(I + y) = \frac{l_d(I)}{EX_1}.$$

**Remarks.** The puzzle is why the particular limit of 10.8 occurs. To get some feel for this, look at the nonlattice case. Let  $N(B)$  denote the number of landings of the sequence  $S_0, S_1, \dots$  in any Borel set  $B$ . Suppose that  $\lim_y EN(B + y)$  existed for all  $B \in \mathcal{B}_1$ . Denote this limit by  $Q(B)$ . Note that  $B_1, B_2$  disjoint implies  $N(B_1 \cup B_2) = N(B_1) + N(B_2)$ . Hence  $Q(B)$  is a nonnegative, additive set function. With a little more work, its  $\sigma$ -additivity and  $\sigma$ -finiteness can be established. The important fact now is that for any  $x \in R^{(1)}$ ,  $B \in \mathcal{B}_1$ ,  $Q(B + x) = Q(B)$ ; hence  $Q$  is invariant under translations. Therefore  $Q$  must be some constant multiple of Lebesgue measure,  $Q(dx) = \alpha dx$ . To get  $\alpha$ , by adding up disjoint intervals, see that

$$EN([0, x]) \sim \alpha x.$$

Let  $N(x) = N([0, x])$ , and assume in addition that  $X_1 \geq 0$  a.s., then

$$N(x) = \max \{n; S_n \leq x\}.$$

By the law of large numbers,

$$(10.9) \quad \frac{S_n}{n} \xrightarrow{\text{a.s.}} EX_1.$$

Along every sample sequence such that (10.9) holds  $N(x) \rightarrow \infty$ , and

$$\lim_{x \rightarrow \infty} \frac{S_{N(x)}}{N(x)} = EX_1, \quad \lim_{x \rightarrow \infty} \frac{S_{N(x)+1}}{N(x) + 1} = EX_1.$$

Since

$$\frac{S_{N(x)}}{N(x)} \leq \frac{x}{N(x)} < \frac{S_{N(x)+1}}{N(x)},$$



then for the sequences such that (10.9) holds,

$$\frac{N(x)}{x} \rightarrow \frac{1}{EX_1}.$$

If we could take expectations of this then  $\alpha = 1/EX_1$ . (This argument is due to Doob [35].)

*Proof of 10.8.* Define the renewal measure on  $\mathcal{B}_1$  by

$$H(B) = EN(B), \quad B \in \mathcal{B}_1.$$

The theorem states that  $H(\cdot + y) \xrightarrow{w} I_d(\cdot)/EX_1$  as  $y$  goes to plus infinity through  $L_d$ . For technical reasons, it is easier to define

$$\mu_v(B) = H(B + y) + H(-B - y), \quad B \in \mathcal{B}_1.$$

The second measure converges weakly to zero as  $y$  goes to plus infinity through  $L_d$ . Because (from Chapter 3, Section 7) if  $I$  is any interval of length  $a$ , then

$$EN(I) \leq P(S_n \in I \text{ at least once}) EN([-a, +a]).$$

So for  $I = [b - a, b]$

$$EN(I - y) \leq P\left(\inf_n S_n \leq b - y\right) EN([-a, +a]).$$

The fact that  $EX_1 > 0$ , implies that the sums are transient, and  $EN(I) < \infty$  for all finite intervals  $I$ . Since  $S_n \rightarrow \infty$  a.s.,  $\inf_n S_n$  is a.s. finite. Hence the right-hand side above goes to zero as  $y \rightarrow +\infty$ . So it is enough to show that  $\mu_v \xrightarrow{w} I_d/EX_1$ . Let  $h \in \mathcal{H}$ . The first order of business is to evaluate  $\int h(x)\mu_v(dx)$ . Since

$$N(B) = \sum_0^\infty \chi_B(S_n), \quad B \in \mathcal{B}_1,$$

then

$$H(B) = \sum_0^\infty P(S_n \in B).$$

Using

$$h(x) = \int e^{iux} \hat{h}(u) du$$

gives

$$\int h(x)H(dx + y) = \sum_0^\infty \iint \hat{h}(u) e^{iux} P(S_n \in dx + y) du,$$

where we assume either that  $h(x)$  is nonnegative, or  $\int |h(x)| \mu_v(dx) < \infty$ . Note that

$$\int e^{iux} P(S_n \in dx + y) = e^{-iuy} \int e^{iux} P(S_n \in dx) = e^{-iuy} [f(u)]^n,$$

so

$$\int h(x)H(dx + y) = \sum_0^\infty \int e^{-iuv} \hat{h}(u) [f(u)]^n du.$$

We would like to take the sum inside the integral sign, but the divergence at  $u = 0$  of  $\sum_0^\infty [f(u)]^n$  is troublesome. Alternatively, let

$$H_r(B) = \sum_0^\infty r^n P(S_n \in B), \quad 0 \leq r < 1.$$

Compute as above, that

$$\int h(x)H_r(dx + y) = \sum_0^\infty r^n \int e^{-iuv} \hat{h}(u) [f(u)]^n du.$$

Now there is no trouble in interchanging to get

$$\int h(x)H_r(dx + y) = \int e^{-iuv} \frac{\hat{h}(u)}{1 - rf(u)} du.$$

In the same way,

$$\int h(x)H_r(-dx - y) = \int e^{-iuv} \frac{\bar{\hat{h}}(u)}{1 - r\bar{f}(u)} du,$$

where  $\bar{f}(u)$  denotes the complex conjugate of  $f(u)$ . Then

$$(10.10) \quad \int h(x)\mu_y(dx) = 2 \lim_{r \uparrow 1} \int e^{-iuv} \hat{h}(u) Rl\left(\frac{1}{1 - rf(u)}\right) du.$$

The basic result used now is a lemma due to Feller and Orey [60].

**Lemma 10.11.** *Let  $f(u)$  be a characteristic function such that  $f(u) \neq 1$  for  $0 < |u| \leq b$ . Then on  $|u| \leq b$ , the measure with density  $Rl(1/1 - rf(u))$  converges weakly as  $r \uparrow 1$  to the measure with density  $Rl(1/1 - f(u))$ , plus a point mass of amount  $\pi/EX_1$  at  $u = 0$ . Also, the integral of  $|Rl(1/1 - f(u))|$  is finite on  $|u| \leq b$ .*

This lemma is really the core of the proof. Accept it for the moment. In the nonlattice case,  $f(u) \neq 1$  for every  $u \neq 0$ . Thus, the limit on the right-hand side of (10.10) exists and

$$\int h(x)\mu_y(dx) = 2\pi \frac{\hat{h}(0)}{EX_1} + 2 \int e^{-iuv} \hat{h}(u) Rl\left(\frac{1}{1 - f(u)}\right) du.$$

Apply the Riemann-Lebesgue lemma to the integral on the right, getting

$$\lim_{y \rightarrow \infty} \int h(x)\mu_y(dx) = \frac{2\pi}{EX_1} \hat{h}(0).$$

The inversion formula (8.40) yields

$$\lim_{y \rightarrow \infty} \int h(x) \mu_y(dx) = \frac{1}{EX_1} \int h(x) dx,$$

which proves the theorem in the nonlattice case. In the lattice case, say  $d = 1$ , put

$$g(u) = \sum_{-\infty}^{+\infty} \hat{h}(u + 2\pi k).$$

Since  $y \in L_1$ , use the notation  $\mu_n$  instead of  $\mu_y$ . Then (10.10) becomes, since  $f(u)$  is periodic,

$$\int h(x) \mu_n(dx) = 2 \lim_{r \uparrow 1} \int_{-\pi}^{+\pi} e^{-iun} g(u) Rl\left(\frac{1}{1 - rf(u)}\right) du.$$

Furthermore, since  $L_1$  is the minimal lattice for  $X_1$ ,  $f(u) \neq 1$  on  $0 < |u| \leq \pi$ . Apply lemma 10.11 again:

$$\int h(x) \mu_n(dx) = 2\pi \frac{g(0)}{EX_1} + 2 \int_{-\pi}^{+\pi} e^{-iun} g(u) Rl\left(\frac{1}{1 - f(u)}\right) du.$$

The Riemann-Lebesgue lemma gives

$$\lim_n \int h(x) \mu_n(dx) = 2\pi \frac{g(0)}{EX_1}.$$

Now look:

$$h(m) = \int e^{iun} \hat{h}(u) du = \int_{-\pi}^{+\pi} e^{iun} g(u) du.$$

By the inversion formula for Fourier series, taking  $\hat{h}(u)$  so that  $\sum_{-\infty}^{+\infty} |h(m)| < \infty$ ,

$$g(u) = \frac{1}{2\pi} \sum_{-\infty}^{+\infty} e^{-iun} h(m),$$

$$g(0) = \frac{1}{2\pi} \int h(x) l_1(dx),$$

finishing the proof in the lattice case.

Now for the proof of the lemma. Let  $\varphi(u)$  be any real-valued continuous function vanishing for  $|u| \geq b$ , such that  $\sup |\varphi| \leq 1$ . Consider

$$\begin{aligned} \int \varphi(u) Rl\left[\frac{1}{1 - rf(u)} - \frac{1}{1 - f(u)}\right] du \\ = - \int \frac{(1 - r)\varphi(u)}{|1 - rf(u)|^2} Rl\left[\frac{f(u)(1 - rf(u))}{1 - f(u)}\right] du. \end{aligned}$$

For any  $\epsilon > 0$ , the integrand converges boundedly to zero on  $(-\epsilon, +\epsilon)^c$  as  $r \uparrow 1$ . Consider the integral over  $(-\epsilon, +\epsilon)$ . The term in brackets equals.

$$(10.12) \quad \frac{f(u)(1 - \bar{f}(u))}{1 - f(u)} + (1 - r) \frac{|f(u)|^2}{1 - f(u)}.$$

Using  $|1 - rf| \geq 1 - r$ , we find that the integral above containing the second term of (10.12) over  $(-\epsilon, +\epsilon)$  is dominated by

$$\int_{|u| < \epsilon} |Rl(1/1 - f)| du.$$

Assuming that  $|Rl(1/1 - f)|$  has a finite integral, we can set this term arbitrarily small by selection of  $\epsilon$ . The function  $(1 - \bar{f})/(1 - f)$  is continuous for  $0 < |u| \leq b$ . Use the Taylor expansion,

$$f(u) = 1 + iu(EX + \delta(u)),$$

where  $\delta(u) \rightarrow 0$  as  $u \rightarrow 0$ , to conclude that the limit of  $(1 - \bar{f})/(1 - f)$  exists equal to  $-1$  as  $u \rightarrow 0$ . Define its value at zero to be  $-1$  making  $(1 - \bar{f})/(1 - f)$  a continuous function on  $|u| \leq b$ . Then the integral containing the first term of (10.12) is given by

$$(10.13) \quad \lim_{r \uparrow 1} \int_{|u| < \epsilon} \frac{1 - r}{|1 - rf(u)|^2} g(u) du,$$

where  $g(u)$  is a continuous function such that  $g(0) = \varphi(0)$ . Denote  $m = EX$ . Use the Taylor expansion again to see that for any  $\Delta > 0$ ,  $\epsilon$  can be selected so that for  $|u| < \epsilon$

$$(1 - \Delta)[(1 - r)^2 + u^2 m^2] \leq |1 - rf(u)|^2 \leq (1 + \Delta)[(1 - r)^2 + u^2 m^2].$$

Combine this with

$$\begin{aligned} \lim_{r \uparrow 1} \int_{|u| < \epsilon} \frac{1 - r}{(1 - r)^2 + u^2 m^2} du &= \lim_{r \uparrow 1} \int_{|v| < \epsilon/(1-r)} \frac{1}{1 + v^2 m^2} dv \\ &= \int_{-\infty}^{+\infty} \frac{1}{1 + v^2 m^2} dv = \frac{\pi}{m} \end{aligned}$$

to conclude that the limit of (10.13) is  $\pi\varphi(0)/m$ . The last fact we need now is the integrability of

$$Rl\left(\frac{1}{1 - f(u)}\right) = \frac{Rl(1 - f(u))}{|1 - f(u)|^2}$$

on  $|u| \leq b$ . Since  $|1 - f(u)|^2 \geq m^2 u^2 / 2$  in some neighborhood of the origin, it is sufficient to show that  $Rl(1 - f(u))/u^2$  has a finite integral. Write

$$\frac{1}{u^2} Rl(f(u) - 1) = \frac{1}{u^2} \int Rl(e^{iux} - 1 - i \sin ux) dF(x),$$

so that

$$\begin{aligned} \int \frac{1}{u^2} R(1 - f(u)) du &\leq \iint \frac{1}{u^2} |e^{iux} - 1 - i \sin ux| dF(x) du \\ &\leq J \int |x| dF(x), \end{aligned}$$

where the orders of integration have been interchanged, and

$$\begin{aligned} J &= \int \frac{1}{u^2 x^2} |e^{iux} - 1 - i \sin ux| |x| du \\ &= \int \frac{1}{v^2} |e^{iv} - 1 - i \sin v| dv < \infty. \end{aligned}$$

#### 4. A LOCAL CENTRAL LIMIT THEOREM

For fair coin-tossing, a combinatorial argument gives  $P(S_{2n} = 0) \sim 1/\sqrt{\pi n}$ , a result that has been useful. What if the  $X_1, X_2, \dots$  are integer-valued, or distributed on the lattice  $L_a$ ? More generally, what about estimating  $P(S_n \in I)$ ? Look at this optimistic argument:

$$(10.14) \quad P(S_n < \sigma \sqrt{nx}) \rightarrow \Phi(x)$$

if  $EX_1 = 0$ ,  $EX_1^2 = \sigma^2 < \infty$ . By Problem 4, Chapter 8, the supremum of the difference in (10.14) goes to zero. Thus by substituting  $y = \sigma \sqrt{n} x$ ,

$$P(y_1 \leq S_n \leq y_2) - \frac{1}{\sqrt{2\pi}} \int_{y_1/\sigma\sqrt{n}}^{y_2/\sigma\sqrt{n}} e^{-x^2/2} dx \rightarrow 0.$$

Substitute in the integral, and rewrite as

$$(10.15) \quad P(y_1 \leq S_n < y_2) - \frac{1}{\sigma\sqrt{2\pi n}} \int_{y_1}^{y_2} e^{-\xi^2/2\sigma^2 n} d\xi \rightarrow 0.$$

If the convergence is so rapid that the difference in (10.15) is  $o(1/\sqrt{n})$ , then multiplying by  $\sigma\sqrt{2\pi n}$  gets us

$$\sigma\sqrt{2\pi n} P(y_1 \leq S_n < y_2) - \int_{y_1}^{y_2} e^{-\xi^2/2\sigma^2 n} d\xi \rightarrow 0.$$

The integrand goes uniformly to one, so rewrite this as

$$\sigma\sqrt{2\pi n} P(y_1 \leq S_n < y_2) \rightarrow y_2 - y_1.$$

This gives the surprising conclusion that estimates like  $P(S_{2n} = 0) \sim 1/\sqrt{\pi n}$  may not be peculiar to fair coin-tossing, but may hold for all sums of independent, identically distributed random variables with zero means and

finite variance. The fact that this delicate result is true for most distributions is a special case of the local central limit theorem. It does not hold generally—look at coin-tossing again:  $P(S_n = 0) = 0$  for  $n$  odd. The next definition restricts attention to those  $X_1, X_2, \dots$  such that the sums do not have an unpleasant periodicity.

**Definition 10.16.** A random variable  $X$  is called a *centered lattice random variable* if there exists  $d > 0$  such that  $P(X \in L_d) = 1$ , and there is no  $d' > d$  and  $\alpha$  such that  $P(X \in \alpha + L_{d'}) = 1$ .  $X$  is called *centered nonlattice* if there are no numbers  $\alpha$  and  $d > 0$  such that  $P(X \in \alpha + L_d) = 1$ .

For example, a random variable  $X$  with  $P(X = 1) = P(X = 3) = \frac{1}{2}$  is not centered lattice, because  $L_1$  is the minimal lattice with  $P(X \in L) = 1$ , but  $P(X \in 1 + L_2) = 1$ .

As before, let  $l_d$  assign mass  $d$  to every point of  $L_d$ , and  $l_0$  denotes Lebesgue measure on  $L_0$ .

**Theorem 10.17.** Let  $X_1, X_2, \dots$  be independent, identically distributed random variables, either centered lattice on  $L_d$  or centered nonlattice on  $L_0$ , with  $EX_1^2 = \sigma^2 < \infty$ . Then for any finite interval  $I$ ,

$$(10.18) \quad \sigma\sqrt{2\pi n}P(S_n \in I) \rightarrow l_d(I).$$

*Proof.* In stages—first of all, if  $X$  is centered nonlattice, then by Problem 21, Chapter 8,  $|f(u)| < 1$ ,  $u \neq 0$ . If  $X$  is centered lattice on  $L_d$ , then  $f(u)$  has period  $2\pi/d$  and the only points at which  $|f(u)| = 1$  are  $\{2\pi k/d\}$ ,  $k = 0, \pm 1, \dots$ . Eq. (10.18) is equivalent to the assertion that the measures  $\mu_n$  defined on  $\mathcal{B}_1$  by

$$\mu_n(B) = \sigma\sqrt{2\pi n}P(S_n \in B),$$

converge weakly to the measure  $l_d$ . The plan of this proof is to show that for every  $h(x) \in \mathcal{K}$

$$(10.19) \quad \sigma\sqrt{2\pi n}Eh(S_n) \rightarrow \int h(x)l_d(dx),$$

and then to apply Theorem 10.7.

Now to prove (10.19): Suppose first that  $|f(u)| \neq 1$  on  $J - \{0\}$ ,  $J$  some finite closed interval, and that  $\hat{h}(u)$  vanishes on  $J^c$ . Write

$$Eh(S_n) = \iint e^{iux} \hat{h}(u) du dF_n(x),$$

where  $F_n$  is the distribution function of  $S_n$ . Then

$$Eh(S_n) = \int [f(u)]^n \hat{h}(u) du.$$

From 8.44,  $f(u) = 1 - (\sigma^2 u^2/2)(1 + \delta(u))$ , where  $\delta(u) \rightarrow 0$  as  $u \rightarrow 0$ . Take  $N = (-b, +b)$ , so small that on  $N$ ,  $|\delta(u)| \leq \frac{1}{2}$ ,  $\sigma^2 u^2 \leq 1$ . On  $J - N$ ,  $|f(u)| \leq 1 - \beta$ ,  $0 < \beta < 1$ . Letting  $\|\hat{h}\| = \sup |\hat{h}(u)|$ , we get

$$(10.20) \quad Eh(S_n) = \int_N [f(u)]^n \hat{h}(u) du + \theta_n \|\hat{h}\| (1 - \beta)^n, \quad |\theta_n| \leq \|J\|.$$

On  $N$ ,

$$(10.21) \quad |f(u)| \leq \left| 1 - \frac{\sigma^2 u^2}{2} \right| + \frac{\sigma^2 u^2}{2} |\delta(u)| \\ \leq 1 - \frac{\sigma^2 u^2}{2} + \frac{\sigma^2 u^2}{4} \leq 1 - \frac{\sigma^2 u^2}{4} \leq e^{-\sigma^2 u^2/4}.$$

By the substitution  $u = v/\sqrt{n}$ ,

$$\int_{\{|u| < b\}} [f(u)]^n \hat{h}(u) du = \int_{\{|v| < b\sqrt{n}\}} \left[ f\left(\frac{v}{\sqrt{n}}\right) \right]^n \hat{h}\left(\frac{v}{\sqrt{n}}\right) \frac{dv}{\sqrt{n}}.$$

By (10.21) the integrand on the right is dominated for all  $v$  by the integrable function

$$\|\hat{h}\| e^{-\sigma^2 v^2/4}.$$

But just as in the central limit theorem,  $[f(v/\sqrt{n})]^n \rightarrow e^{-\sigma^2 v^2/2}$ . Since  $\hat{h}(v/\sqrt{n}) \rightarrow \hat{h}(0)$ , use the dominated convergence theorem for

$$\sqrt{n} \int_N [f(u)]^n \hat{h}(u) du \rightarrow \hat{h}(0) \int e^{-\sigma^2 v^2/2} dv = \hat{h}(0) \frac{\sqrt{2\pi}}{\sigma}.$$

Use (10.20) to get

$$(10.22) \quad \sqrt{n} Eh(S_n) \rightarrow \frac{\sqrt{2\pi}}{\sigma} \hat{h}(0).$$

When  $X$  is centered nonlattice, this holds for all finite  $J$ . Furthermore, the Fourier inversion theorem gives

$$\hat{h}(u) = \frac{1}{2\pi} \int e^{-iux} h(x) dx.$$

By putting  $u = 0$  we can prove the assertion. In the lattice case, assume  $d = 1$  for convenience, so  $\hat{f}(u)$  has period  $2\pi$ . Let  $g(u) = \sum_{-\infty}^{+\infty} \hat{h}(u + 2\pi k)$  so that

$$Eh(S_n) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} [f(u)]^n g(u) du.$$

The purpose of this is that now  $|f(u)| \neq 1$  on  $[-\pi, +\pi] - \{0\}$ , so (10.22) holds in the form

$$\sqrt{n} Eh(S_n) \rightarrow \frac{\sqrt{2\pi}}{\sigma} g(0).$$

Just as in the proof of the renewal theorem,

$$g(0) = \frac{1}{2\pi} \int h(x) l_1(dx)$$

which proves the theorem in the lattice case. Q.E.D.

**Problem 1.** Under the conditions of this section, show that

$$\sigma \sqrt{2\pi n} P(S_n \in I + x) \rightarrow l_d(I)$$

uniformly for  $x$  in bounded subsets of  $L_d$ . [See Problem 4, Chapter 8.]

## 5. APPLYING A TAUBERIAN THEOREM

Let  $X_1, X_2, \dots$  be centered lattice. Suppose we want to get information concerning the distribution of  $T$ , the time of the first zero of the sums  $S_1, S_2, \dots$ . From (7.42), for all  $n \geq 0$ ,

$$1 = \sum_{m=0}^n P(S_{n-m} = 0) P(T > m)$$

with the convention  $S_0 \equiv 0$ . Multiply this equation by  $r^n$ ,  $0 \leq r < 1$ , and sum from  $n = 0$  to  $\infty$ ,

$$\frac{1}{1-r} = \left( \sum_0^\infty r^n P(S_n = 0) \right) \left( \sum_0^\infty r^n P(T > n) \right).$$

The local limit theorem gives

$$P(S_n = 0) \sim \frac{d}{\sigma \sqrt{2\pi n}},$$

so as  $r \uparrow 1$ ,

$$P(r) = \sum_0^\infty r^n P(S_n = 0)$$

blows up. Suppose we can use the asymptotic expression for  $P(S_n = 0)$  to get the rate at which  $P(r)$  blows up. Since

$$T(r) = \sum_0^\infty r^n P(T > n)$$

is given by

$$T(r) = \frac{1}{(1-r)P(r)},$$

we have information regarding the rate at which  $T(r) \rightarrow \infty$  when  $r \rightarrow 1$ . Now we would like to reverse direction and get information about  $P(T > n)$  from the rate at which  $T(r)$  blows up. The first direction is called an Abelian argument; the reverse and much more difficult direction is called a Tauberian argument.



To get some feeling for what is going on, consider

$$\varphi(r) = \sum_1^{\infty} n^{\alpha} r^n, \quad \alpha > -1.$$

Put  $r = e^{-s}$ . For  $s$  small, a good approximation to  $\varphi(r)$  is given by the integral

$$\int_0^{\infty} e^{-sx} x^{\alpha} dx = \frac{1}{s^{\alpha+1}} \int_0^{\infty} e^{-x} x^{\alpha} dx.$$

The integral is the gamma function  $\Gamma(\alpha + 1)$ . Since  $s \sim 1 - r$  as  $r \uparrow 1$ , this can be made to provide a rigorous proof of

$$(10.23) \quad \varphi(r) \sim \frac{\Gamma(\alpha + 1)}{(1 - r)^{\alpha+1}}, \quad \text{as } r \uparrow 1.$$

Now suppose (10.23) holds for  $\varphi(r) = \sum_0^{\infty} a_n r^n$ . Can we reverse and conclude that  $a_n \sim n^{\alpha}$ ? Not quite; what is true is the well-known theorem:

**Theorem 10.24.** Let  $\varphi(r) = \sum_0^{\infty} a_n r^n$ ,  $a_n \geq 0$ ,  $n = 0, \dots$ . Then as  $n \rightarrow \infty$

$$a_0 + \dots + a_n \sim cn^{\alpha}, \quad \alpha > 0,$$

if and only if, as  $r \uparrow 1$ ,

$$\varphi(r) \sim c(1 - r)^{-\alpha} \Gamma(\alpha + 1).$$

For a nice proof of this theorem, see Feller [59, Vol. II, p. 423]. The implication from the behavior of  $\varphi(r)$  to that of  $a_1 + \dots + a_n$  is the hard part and is a special case of Karamata's Tauberian theorem [85].

We use the easy part of this theorem to show that

$$P(r) \sim (1 - r)^{-1/2} \frac{d}{\sigma\sqrt{2}}.$$

This follows immediately from the local limit theorem, the fact that  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ , and

$$\sum_1^n \frac{1}{\sqrt{k}} \sim 2\sqrt{n}.$$

[Approximate the sums above and below by integrals of  $1/\sqrt{x}$  and  $1/\sqrt{x+1}$  respectively.]

Of course, this gives

$$T(r) \sim \frac{\sigma\sqrt{2}}{d} (1 - r)^{-1/2}, \quad \text{as } r \uparrow 1.$$

**Theorem 10.25**

$$P(T > n) \sim \frac{\sigma}{d} \frac{\sqrt{2}}{\sqrt{n\pi}}.$$

*Proof.* By Karamata's Tauberian theorem, putting  $p_n = P(T > n)$ ,  $c = 2^{3/2}\sigma/d\sqrt{\pi}$ , we get

$$p_0 + \cdots + p_n \sim c\sqrt{n}.$$

Since the  $p_n$  are nonincreasing, for any  $m \leq n$  write

$$(n - m)p_n \leq p_{m+1} + \cdots + p_n \leq (n - m)p_{m+1}.$$

Divide by  $\sqrt{n}$ , put  $m = [\lambda n]$ ,  $0 < \lambda < 1$ , and let  $n \rightarrow \infty$  to get

$$(1 - \lambda) \lim \sqrt{n} p_n \leq c(1 - \sqrt{\lambda}) \leq \frac{(1 - \lambda)}{\sqrt{\lambda}} \lim \sqrt{m} p_m.$$

Let  $\lambda \rightarrow 1$ ; then  $(1 - \sqrt{\lambda})/(1 - \lambda) \rightarrow \frac{1}{2}$ , so  $\sqrt{n} p_n \rightarrow c/2$ .

**Problem 2.** Let  $N_n(0)$  be the number of times that the sums  $S_1, \dots, S_n$  visit the state zero. Use the theory of stable laws and 10.25 to find constants  $A_n \uparrow \infty$  such that  $N_n(0)/A_n \xrightarrow{\mathcal{D}} X$  where  $X$  is nondegenerate.

**6. OCCUPATION TIMES**

One neat application of the local central limit theorem is to the problem: Given a finite interval  $I$ ,  $S_1, S_2, \dots$  sums of independent, identically distributed random variables  $X_1, X_2, \dots$  such that  $EX_1 = 0$ ,  $EX_1^2 = \sigma^2 < \infty$ , and take  $S_0 \equiv 0$ . Let  $N_n$  be the number of visits of  $S_0, S_1, \dots, S_n$  to the interval  $I$ . Is there a normalization such that  $N_n/A_n \xrightarrow{\mathcal{D}} X$ ,  $X$  nondegenerate?  $N_n$  is the amount of time in the first  $n$  trials that the sums spend in the interval  $I$ , hence the name "occupation-time problem." Let  $X_1, X_2, \dots$  be centered lattice or nonlattice, so the local limit theorem is in force. To get the size of  $A_n$ , compute

$$EN_n = \sum_0^n E\chi_I(S_k) \sim \sum_0^n \frac{l_d(I)}{\sigma\sqrt{2\pi n}} \sim c\sqrt{n}.$$

Hence we take  $A_n = \sqrt{n}$ . The way from here is the method of moments. Let

$$\mu_k(n) = E(N_n)^k = \sum_{n \geq j_k, \dots, j_1 \geq 0} E(\chi_I(S_{j_k}) \cdots \chi_I(S_{j_1})).$$

By combining all permutations of the same indices  $j_1, \dots, j_k$  we get

$$\mu_k(n) = k! \sum_{n \geq j_k \geq \dots \geq j_1 \geq 0} E(\chi_I(S_{j_k}) \cdots \chi_I(S_{j_1}))$$

and

$$\mu_k(n) - \mu_k(n-1) = k! \sum_{n \geq j_{k-1} \geq \dots \geq j_1 \geq 0} E(\chi_I(S_n) \chi_I(S_{j_{k-1}}) \cdots \chi_I(S_{j_1})).$$

Define transition probabilities on  $\mathcal{B}_1$  by

$$p^{(j)}(B | x) = P(S_{k+j} \in B | S_k = x) = P(S_j \in B - x).$$

So

$$\mu_k(n) - \mu_k(n-1) = k! \sum_{\substack{i_1, \dots, i_k \geq 0, \\ i_1 + \dots + i_k = n}} \int_I \cdots \int_I \prod_{m=1}^k p^{(i_m)}(dx_m | x_{m-1}),$$

where  $x_0 \equiv 0$ . If

$$p_r(B | x) = \sum_0^\infty r^n p^{(n)}(B | x)$$

where  $p^{(0)}(B | x)$  is the point mass concentrated on  $\{x\}$ , and  $0 \leq r < 1$ , then, defining  $\mu_k(-1) = 0$ ,

$$(10.26) \quad \sum_{n=0}^\infty r^n (\mu_k(n) - \mu_k(n-1)) = k! \int_I \cdots \int_I \prod_{m=1}^k p_r(dx_m | x_{m-1}).$$

**Proposition 10.27**

$$\sqrt{1-r} p_r(I | x) \xrightarrow{uc} \frac{1}{\sigma\sqrt{2}} l_d(I)$$

for  $x \in L_d$  as  $r \uparrow 1$ .

*Proof.* By Problem 1 of this chapter

$$\sigma \sqrt{2\pi n} p^{(n)}(I | x) \xrightarrow{uc} l_d(I)$$

for  $x \in L_d$ . Hence for all  $x \in J \cap L_d$ ,  $J$  a finite interval, and for any  $\epsilon > 0$ , there exists  $n_0$  such that for  $n \geq n_0$ ,

$$\frac{(1-\epsilon)l_d(I)}{\sigma\sqrt{2\pi n}} \leq p^{(n)}(I | x) \leq \frac{(1+\epsilon)l_d(I)}{\sigma\sqrt{2\pi n}}.$$

By 10.24 we get *uc* convergence of  $\sqrt{1-r} p_r(I | x)$  to  $2\Gamma(\frac{3}{2})l_d(I)/\sigma\sqrt{2\pi}$ .

**Proposition 10.28**

$$\lim_{r \uparrow 1} (1-r)^{k/2} \sum_{n=0}^\infty r^n (\mu_k(n) - \mu_k(n-1)) = k! \left( \frac{l_d(I)}{\sigma\sqrt{2}} \right)^k.$$

*Proof.* On the right-hand side of (10.26), look at the first factor of the integrand multiplied by  $\sqrt{1-r}$ , namely

$$\sqrt{1-r} p_r(I | x_{k-1}).$$

This converges uniformly for  $x_{k-1} \in I \cap L_d$  to a constant. Hence we can pull it out of the integral. Now continue this way to get the result above.

With 10.28 in hand, apply Karamata's Tauberian theorem to conclude that

$$(10.29) \quad \mu_k(n) \sim \frac{n^{k/2}}{\Gamma((k/2) + 1)} k! \left( \frac{l_d(I)}{\sigma\sqrt{2}} \right)^k.$$

This almost completes the proof of

**Theorem 10.30**

$$\frac{N_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \frac{l_d(I)}{\sigma\sqrt{2}} X,$$

where

$$P(X < x) = \frac{1}{\sqrt{\pi}} \int_0^x e^{-t^2/4} dt.$$

*Proof.* By (10.29),

$$E\left(\frac{N_n}{\sqrt{n}}\right)^k \rightarrow \frac{k!}{\Gamma((k/2) + 1)} \left(\frac{l_d(I)}{\sigma\sqrt{2}}\right)^k.$$

Let

$$m_k = EX^k.$$

Use integration by parts to show that

$$m_{k+2} = 2(k+1)m_k,$$

and deduce from this that

$$\frac{k!}{\Gamma((k/2) + 1)} = EX^k.$$

Proposition 8.49 implies that the distribution is here uniquely determined by the moments. Therefore Theorem 8.48 applies. Q.E.D.

## NOTES

The renewal theorem in the lattice case was stated and proved by Erdős, Feller, and Pollard [49] in 1949. Chung later pointed out that in this case the theorem follows from the convergence theorem for countable Markov chains due to Kolmogorov (see Chapter 7, Section 7). The theorem was gradually extended and in its present form was proved by Blackwell [6, 7]. New proofs continue to appear. There is an interesting recent proof due to Feller, see [59, Vol. II], which opens a new aspect of the theorem. The method of proof we use is adapted from Feller and Orey [60, 1961]. Charles Stone [132] by similar methods gets very accurate estimates for the rate of convergence of  $H(B+x)$  as  $x \rightarrow \infty$ . A good exposition of the state of renewal theory as of 1958 is given by W. L. Smith [127]. One more result that is usually considered to be part of the renewal theorem concerns the case  $EX = \infty$ . What is known here is that if the sums  $S_0, S_1, \dots$  are transient, then  $H(I+y) \rightarrow 0$  as  $y \rightarrow \pm\infty$  for all finite intervals  $I$ . Hence, in particular, if one of  $EX^+$ ,  $EX^-$  is finite, this result holds (see Feller [59, Vol. II, pp. 368 ff]).

Local limit theorems for lattice random variables have a long history. The original proof of the central limit theorem for coin-tossing was gotten by first estimating  $P(S_n = j)$  and thus used a local limit theorem to prove the

tendency toward  $\mathcal{N}(0, 1)$ . More local limit theorems for the lattice case are in Gnedenko and Kolmogorov [62]. Essentially the theorem given in the text is due to Shepp [122]; the method of proof follows [133]. In its form in the text, the central limit theorem is not a consequence of the local theorem. But by very similar methods somewhat sharper results can be proved. For example, in the centered nonlattice case, for any interval  $I$ , let

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}.$$

Stone proves that

$$\sup_x |\sqrt{n}P(S_n \in I + x) - l(I)\varphi(x/\sqrt{n})| \rightarrow 0,$$

and the central limit theorem for centered nonlattice variables follows from this.

The occupation time theorem 10.30 was proven for normally distributed random variables by Chung and Kac [19], and in general, by Kallianpur and Robbins [84]. Darling and Kac [23] generalized their results significantly and simplified the proof by adding the Tauberian argument.

The occupation time problem for an infinite interval, say  $(0, \infty)$ , is considerably different. Then  $N_n$  becomes the number of positive sums among  $S_1, \dots, S_n$ . The appropriate normalizing factor is  $n$ , and the famous arc sine theorem states

$$P\left(\frac{N_n}{n} < x\right) \rightarrow \frac{2}{\pi} \arcsin \sqrt{x}.$$

See Spitzer's book [130], Chapter 4, for a complete discussion in the lattice case, or Feller [59, Vol. II, Chapter 12], for the general case.

## CHAPTER 11

# MULTIDIMENSIONAL CENTRAL LIMIT THEOREM AND GAUSSIAN PROCESSES

### 1. INTRODUCTION

Suppose that the objects under study are a sequence  $\mathbf{X}_n$  of vector-valued variables  $(X_1^{(n)}, \dots, X_k^{(n)})$  where each  $X_j^{(n)}$ ,  $j = 1, \dots, k$ , is a random variable. What is a reasonable meaning to attach to

$$\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X},$$

where  $\mathbf{X} = (X_1, \dots, X_k)$ ? Intuitively, the meaning of convergence in distribution is that the probability that  $\mathbf{X}_n$  is in some set  $B \in \mathcal{B}_k$  converges to the probability that  $\mathbf{X}$  is in  $B$ , that is,

$$P(\mathbf{X}_n \in B) \rightarrow P(\mathbf{X} \in B).$$

But when we attempt to make this hold for all  $B \in \mathcal{B}_k$ , difficulty is encountered in that  $\mathbf{X}$  may be degenerate in part. In the one-dimensional case, the one-point sets to which the limit  $\mathbf{X}$  assigned positive probability gave trouble and had to be excluded. In general, what can be done is to require that

$$P(\mathbf{X}_n \in B) \rightarrow P(\mathbf{X} \in B)$$

for all sets  $B \in \mathcal{B}_k$  such that  $P(\mathbf{X} \in \text{bd}(B)) = 0$ .

The definition we use is directed at the problem from a different but equivalent angle. Let  $\mathcal{E}_0$  be the class of all continuous functions on  $R^{(k)}$  vanishing off of compact sets.

**Definition 11.1.** The  $k$ -vectors  $\mathbf{X}_n$  converge in distribution (or in law) to  $\mathbf{X}$  if for every  $f(\mathbf{x}) \in \mathcal{E}_0$ ,

$$Ef(\mathbf{X}_n) \rightarrow Ef(\mathbf{X}).$$

This is written as  $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}$  or  $\mathcal{L}(\mathbf{X}_n) \rightarrow \mathcal{L}(\mathbf{X})$ . In terms of distribution functions,  $F_n \xrightarrow{\mathcal{D}} F$  if and only if

$$\int f(\mathbf{x})F_n(d\mathbf{x}) \rightarrow \int f(\mathbf{x})F(d\mathbf{x})$$

for all  $f \in \mathcal{E}_0$ .

By considering continuous functions equal to one on some compact set and vanishing on the complement of some slightly larger compact set conclude that if  $F_n \xrightarrow{\mathcal{D}} F$ , and  $F_n \xrightarrow{\mathcal{D}} G$ , then  $F = G$ . Define as in the one-dimensional case:

**Definition 11.2.** Let  $\mathcal{N}_k$  be the set of all distribution functions on  $R^{(k)}$ . A set  $\mathcal{L} \subset \mathcal{N}_k$  is mass-preserving if for any  $\epsilon > 0$ , there is a compact set  $A$  such that  $F(A^c) \leq \epsilon$ , for all  $F \in \mathcal{L}$ .

If  $F_n \xrightarrow{\mathcal{D}} F$ , then  $\{F_n\}$  is mass-preserving. From this, conclude that  $F_n \xrightarrow{\mathcal{D}} F$  if and only if for every bounded continuous function  $f(\mathbf{x})$  on  $R^{(k)}$ ,

$$\int f(\mathbf{x}) F_n(d\mathbf{x}) \rightarrow \int f(\mathbf{x}) F(d\mathbf{x}).$$

For any rectangle  $S$  such that  $F(bd(S)) = 0$ , approximate  $\chi_S$  above and below by continuous functions to see that  $F_n(S) \rightarrow F(S)$ . Conversely, approximate  $\int f dF_n$  by Riemann sums over rectangles such that  $F(bd(S)) = 0$  to conclude that  $F_n \xrightarrow{\mathcal{D}} F$  is equivalent to

$$F_n(S) \rightarrow F(S), \quad \text{if} \quad F(bd(S)) = 0.$$

There are plenty of rectangles whose boundaries have probability zero, because if  $P(\mathbf{X} \in S) = P(\mathbf{X}_1 \in I_1, \dots, \mathbf{X}_k \in I_k)$ , then

$$P(\mathbf{X} \in bd(S)) \leq \sum_j P(\mathbf{X}_j \in bd(I_j)).$$

By the same approximation as in 8.12, conclude that for  $f(\mathbf{x})$  bounded,  $\mathcal{B}_k$ -measurable and with its discontinuity set having  $F$ -measure zero, that

$$\int f dF_n \rightarrow \int f dF.$$

From this, it follows that if  $B$  is in  $\mathcal{B}_k$  and  $F(bd(B)) = 0$ , then  $F_n(B) \rightarrow F(B)$ .

**Problem 1.** Is  $F_n \xrightarrow{\mathcal{D}} F$  equivalent to requiring

$$F_n(x_1, \dots, x_k) \rightarrow F(x_1, \dots, x_k)$$

at every continuity point  $(x_1, \dots, x_k)$  of  $F$ ? Prove or disprove.

## 2. PROPERTIES OF $\mathcal{N}_k$

The properties of  $k$ -dimensional probability measures are very similar to those of one-dimensional probabilities. The results are straightforward generalizations, and we deal with them sketchily. The major result is the generalization of the Helly-Bray theorem.

**Theorem 11.3.** Let  $\{F_n\} \subset \mathcal{N}_k$  be mass-preserving. Then there exists a subsequence  $F_{n_m}$  converging in distribution to some  $F \in \mathcal{N}_k$ .

*Proof.* Here is a slightly different proof that opens the way for generalizations. Take  $\{f_j\} \subset \mathcal{E}_0$  to be dense in  $\mathcal{E}_0$  in the sense of uniform convergence. To verify that a countable set can be gotten with this property, look at the set of all polynomials with rational coefficients, then consider the set gotten by multiplying each of these by a function  $h_N$  which is one inside the  $k$ -sphere of radius  $N$  and zero outside the sphere of radius  $N + 1$ . Use diagonalization again; for every  $j$ , let  $I_j$  be an ordered subset of the positive integers such that  $\int f_j dF_n$  converges as  $n$  runs through  $I_j$ , and  $I_{j+1} \subset I_j$ . Let  $n_m$  be the  $m$ th member of  $I_m$ ; then the limit of  $\int f_j dF_{n_m}$  exists for all  $j$ . Take  $f \in \mathcal{E}_0$ ,  $|f - f_j| \leq \epsilon$  so that

$$\left| \int f dF_{n_m} - \int f_j dF_{n_m} \right| \leq \epsilon,$$

hence  $\lim \int f dF_{n_m}$  exists for all  $f \in \mathcal{E}_0$ . Because  $\{F_n\}$  is mass-preserving,  $\lim \int f dF_{n_m} = J(f)$  exists for all bounded continuous  $f$ . Denote the open rectangle  $\{(x_1, \dots, x_k); x_1 < y_1, \dots, x_k < y_k\}$  by  $S(y)$ . Take  $g_n$  bounded and continuous such that  $g_n \uparrow \chi_{S(y)}$ . Then  $J(g_n)$  is nondecreasing; call the limit  $F(y_1, \dots, y_k)$ . The rest of the proof is the simple verification that  $F(y_1, \dots, y_k)$  is a distribution function, and that if  $F(bd(S(y))) = 0$ , then  $F_n(S(y)) \rightarrow F(S(y))$ .

Define sets  $\mathcal{E}$  of  $\mathcal{N}_k$ -separating functions as in 8.14, prove as in 8.15 and 8.16 that  $\{F_n\}$  mass-preserving,  $\int f dF_n$  convergent, all  $f \in \mathcal{E}$  imply the existence of an  $F \in \mathcal{N}_k$  such that  $F_n \xrightarrow{\mathcal{D}} F$ . Obviously, one sufficient condition for a class of functions to be  $\mathcal{N}_k$ -separating is that they be dense in  $\mathcal{E}_0$  under  $uc$  convergence of uniformly-bounded sequences of functions.

**Theorem 11.4.** *The set of complex exponentials of the form*

$$\{\exp [i(u_1 x_1 + \dots + u_k x_k)]\}, \quad \mathbf{u} \in R^{(k)}$$

*is  $\mathcal{N}_k$ -separating.*

*Proof.* The point here is that for any  $f \in \mathcal{E}_0$ , we can take  $n$  so large that  $f(\mathbf{x}) = 0$  on the complement of  $S_n = \{\mathbf{x}; x_j \in [-n, +n], j = 1, \dots, k\}$ . Now approximate  $f(\mathbf{x})$  uniformly on  $S_n$  by sums of terms of the form  $\exp [\pi i(m_1 x_1 + \dots + m_k x_k)/n]$ ,  $m_1, \dots, m_k$ , integers. The rest goes through as in Theorem 8.24.

**Definition 11.5.** For  $\mathbf{u}, \mathbf{x} \in R^{(k)}$ , let  $(\mathbf{u}, \mathbf{x}) = \sum_1^k u_i x_i$ , and define the characteristic function of the  $k$ -vector  $\mathbf{X} = (X_1, \dots, X_k)$  as  $f_{\mathbf{X}}(\mathbf{u}) = Ee^{i(\mathbf{u}, \mathbf{X})}$  or of the distribution function  $F(\mathbf{x})$  as

$$f(\mathbf{u}) = \int e^{i(\mathbf{u}, \mathbf{x})} F(d\mathbf{x}).$$

The continuity theorem holds.



**Theorem 11.6.** For  $F_n \in \mathcal{N}_k$  having characteristic functions  $f_n(\mathbf{u})$ , if

- a)  $\lim_n f_n(\mathbf{u})$  exists for every  $\mathbf{u} \in R^{(k)}$ ,  
 b)  $\lim_n f_n(\mathbf{u}) = h(\mathbf{u})$  is continuous at the origin,

then there is a distribution function  $F \in \mathcal{N}_k$  such that  $F_n \xrightarrow{\mathcal{D}} F$  and  $h(\mathbf{u})$  is the characteristic function of  $F$ .

*Proof.* The only question at all is the analog of inequality 8.29. But this is simple. Observe that

$$\begin{aligned} \int \frac{1}{u^k} \left[ \int_0^u \cdots \int_0^u \prod_1^k (1 - \cos v_j x_j) dv \right] F(d\mathbf{x}) \\ = \int \prod_1^k \left( 1 - \frac{\sin ux_j}{ux_j} \right) F(d\mathbf{x}) \geq \frac{1}{\alpha^k} F(S_u^c), \end{aligned}$$

where  $S_u = \{|x_1| \leq 1/u, \dots, |x_k| \leq 1/u\}$ . For any function  $g(v_1, \dots, v_k)$  on  $R^{(k)}$  define  $T_j g$  to be the function

$$\begin{aligned} g(v_1, \dots, v_{j-1}, 0, v_{j+1}, \dots, v_k) - \frac{1}{2} g(v_1, \dots, v_j, \dots, v_k) \\ - \frac{1}{2} g(v_1, \dots, -v_j, \dots, v_k). \end{aligned}$$

Then

$$\int \prod_1^k (1 - \cos v_j x_j) F(d\mathbf{x}) = T_k \cdots T_1 f(v_1, \dots, v_k),$$

where  $f$  is the characteristic function of  $F$ . The function  $T_k \cdots T_1 f(\mathbf{v})$  is continuous and is zero at the origin. Write the inequality as

$$F(S_u^c) \leq \alpha^k \cdot \frac{1}{u^n} \int_0^u \cdots \int_0^u T_k \cdots T_1 f(\mathbf{v}) d\mathbf{v}.$$

Now,  $f_n(\mathbf{u}) \rightarrow h(\mathbf{u})$  implies

$$\overline{\lim} F_n(S_u^c) \leq \alpha^k \cdot \frac{1}{u^n} \int_0^u \cdots \int_0^u T_k \cdots T_1 h(\mathbf{v}) d\mathbf{v},$$

and everything goes through.

### Problems

2. If  $\mathbf{X}_1^{(n)}, \dots, \mathbf{X}_k^{(n)}$  are independent for every  $n$ , and if for  $j$  fixed,

$$\mathbf{X}_j^{(n)} \xrightarrow{\mathcal{D}} \mathbf{X}_j,$$

prove that

$$(\mathbf{X}_1^{(n)}, \dots, \mathbf{X}_k^{(n)}) \xrightarrow{\mathcal{D}} (\mathbf{X}_1, \dots, \mathbf{X}_k)$$

and  $\mathbf{X}_1, \dots, \mathbf{X}_k$  are independent.

3. Let  $\mathbf{X}^{(n)} \xrightarrow{\mathcal{D}} \mathbf{X}$  and  $\varphi_1, \dots, \varphi_m$  be continuous functions on  $R^{(k)}$ . Show that

$$(\varphi_1(\mathbf{X}^{(n)}), \dots, \varphi_m(\mathbf{X}^{(n)})) \xrightarrow{\mathcal{D}} (\varphi_1(\mathbf{X}), \dots, \varphi_m(\mathbf{X})).$$

4. Show that the conclusion of Problem 3 remains true if  $\varphi_k, k = 1, \dots, m$ , continuous is replaced by  $\varphi_k(\mathbf{x}), k = 1, \dots, m$  a.s. continuous with respect to the distribution on  $R^{(k)}$  given by  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_k)$ .

### 3. THE MULTIDIMENSIONAL CENTRAL LIMIT THEOREM

To use the continuity theorem, estimates on  $f(\mathbf{u})$  are needed. Write

$$\|\mathbf{x}\| = \sqrt{\sum_1^k x_j^2}.$$

**Proposition 11.7.** Let  $E \|\mathbf{X}\|^n < \infty$ , then

$$f(\mathbf{u}) = \sum_{m=0}^n \frac{i^m}{m!} E[(\mathbf{u}, \mathbf{X})^m] + \delta(\mathbf{u}) \|\mathbf{u}\|^n,$$

where  $\delta(\mathbf{u}) \rightarrow 0$  as  $\mathbf{u} \rightarrow 0$ .

*Proof.* Write

$$e^{i(\mathbf{u}, \mathbf{x})} = \sum_j \frac{i^n}{n!} (\mathbf{u}, \mathbf{x})^n + \frac{(\mathbf{u}, \mathbf{x})^n}{n!} \varphi(\mathbf{u}, \mathbf{x}),$$

where

$$\varphi(\mathbf{u}, \mathbf{x}) = \cos [\theta_1(\mathbf{u}, \mathbf{x})] + i \sin [\theta_2(\mathbf{u}, \mathbf{x})] - 1,$$

$\theta_1, \theta_2$  real and  $|\theta_1| \leq 1, |\theta_2| \leq 1$ . By the Schwarz inequality we have  $|(\mathbf{u}, \mathbf{x})| \leq \|\mathbf{u}\| \|\mathbf{x}\|$ . Thus

$$|E[(\mathbf{u}, \mathbf{X})^n \varphi(\mathbf{u}, \mathbf{X})]| \leq \|\mathbf{u}\|^n E(\|\mathbf{X}\|^n |\varphi(\mathbf{u}, \mathbf{X})|).$$

The integrand is dominated by the integrable function  $3 \|\mathbf{X}\|^n$ , and  $\varphi(\mathbf{u}, \mathbf{X}) \rightarrow 0$  as  $\mathbf{u} \rightarrow 0$  for all  $\omega$ . Apply the bounded convergence theorem to get the result.

**Definition 11.8.** Given a  $k$ -vector  $(\mathbf{X}_1, \dots, \mathbf{X}_k)$ ,  $E\mathbf{X}_j = 0, j = 1, \dots, k$ . Define the  $k \times k$  covariance matrix  $\Gamma$  by

$$\Gamma_{ij} = E\mathbf{X}_i\mathbf{X}_j, \quad i, j = 1, \dots, k.$$

**Definition 11.9.** The vector  $\mathbf{X}, E\mathbf{X}_j = 0, j = 1, \dots, k$ , is said to have a joint normal distribution  $\mathcal{N}(0, \Gamma)$  if

$$f_{\mathbf{X}}(\mathbf{u}) = \exp \left[ -\frac{1}{2} \sum_{i,j} \Gamma_{ij} u_i u_j \right].$$

**Theorem 11.10.** Let  $\mathbf{X}_1, \mathbf{X}_2, \dots$  be independent  $k$ -vectors having the same distribution with zero means and finite covariance matrix  $\Gamma$ . Then

$$\frac{\mathbf{X}_1 + \dots + \mathbf{X}_n}{\sqrt{n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma).$$

*Proof*

$$E(\exp i[(\mathbf{u}, \mathbf{X}_1 + \dots + \mathbf{X}_n)/\sqrt{n}]) = [E \exp i(\mathbf{u}, \mathbf{X})/\sqrt{n}]^n,$$

where  $\mathbf{X}$  has the same distribution as  $\mathbf{X}_1, \mathbf{X}_2, \dots$ . By assumption,  $E\mathbf{X}_j^2 < \infty$ ,  $j = 1, \dots, k$ , where  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_k)$ . Thus  $E \|\mathbf{X}\|^2 < \infty$ . By Proposition 11.7,

$$f(\mathbf{u}) = 1 - \frac{1}{2}E[(\mathbf{u}, \mathbf{X})^2] + \delta(\mathbf{u}) \|\mathbf{u}\|^2,$$

so

$$E e^{i(\mathbf{u}, \mathbf{X})/\sqrt{n}} = f(\mathbf{u}/\sqrt{n}) = 1 - (1/2n)E[(\mathbf{u}, \mathbf{X})^2] + (1/n) \delta(\mathbf{u}/\sqrt{n}) \|\mathbf{u}\|^2.$$

$$\begin{aligned} (f(\mathbf{u}/\sqrt{n}))^n &= \left[ 1 - \frac{1}{n} \left( \frac{1}{2}E[(\mathbf{u}, \mathbf{X})^2] + \delta(\mathbf{u}/\sqrt{n}) \|\mathbf{u}\|^2 \right) \right]^n \\ &\rightarrow e^{-1/2E[(\mathbf{u}, \mathbf{X})^2]} = e^{-1/2 \sum \Gamma_{ij} u_i u_j}. \end{aligned}$$

Once more, as long as second moments exist, the limit is normal. There are analogs here of Theorem 9.2 for the nonidentically distributed case which involve bounds on  $E \|\mathbf{X}_n\|^3$ . But we leave this; the tools are available for as many of these theorems as we want to prove.

#### 4. THE JOINT NORMAL DISTRIBUTION

The neatest way of defining a joint normal distribution is

**Definition 11.11.** Say that  $\mathbf{Y} = (Y_1, \dots, Y_n)$ ,  $EY_i = 0$ , has a joint normal (or joint Gaussian) distribution with zero means if there are  $k$  independent random variables  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_k)$ , each with  $\mathcal{N}(0, 1)$  distribution, and a  $k \times n$  matrix  $A$  such that

$$\mathbf{Y} = \mathbf{X}A.$$

Obviously, the matrix  $A$  and vector  $\mathbf{X}$  are not unique. Say that a set  $Z_1, \dots, Z_j$  of random variables is linearly independent if there are no real numbers  $\alpha_1, \dots, \alpha_j$  not all zero, such that

$$\alpha_1 Z_1 + \dots + \alpha_j Z_j = 0 \quad \text{a.s.}$$

Then note that the minimal  $k$ , such that there is an  $A$ ,  $\mathbf{X}$  as in 11.11, with  $\mathbf{Y} = \mathbf{X}A$ , is the maximum number of linearly independent random variables in the set  $Y_1, \dots, Y_n$ . If

$$\sum_{j=1}^n \alpha_j Y_j = 0 \quad \text{a.s.,}$$

then

$$\sum_{j=1}^n \alpha_j \Gamma_{ij} = 0, \quad i = 1, \dots, n,$$

so the minimum  $k$  is also given by the rank of the covariance matrix  $\Gamma$ . Throughout this section, take the joint normal distribution with zero means to be defined by 11.11. We will show that it is equivalent to 11.9.

If  $\mathbf{Y} = \mathbf{XA}$ , then the covariance matrix of  $\mathbf{Y}$  is given by

$$E\mathbf{Y}_i\mathbf{Y}_j = E\left(\sum_k \mathbf{X}_k a_{kj}\right)\left(\sum_m \mathbf{X}_m a_{mi}\right) = \sum_k a_{ki}a_{kj}.$$

So

$$\Gamma = A^t A.$$

This is characteristic of covariance matrices.

**Definition 11.12.** Call a square matrix  $M$  symmetric if  $m_{ij} = m_{ji}$ , nonnegative definite if  $\alpha M \alpha^t \geq 0$  for all vectors  $\alpha$ , where  $\alpha^t$  denotes the transpose of  $\alpha$ .

**Proposition 11.13.** An  $n \times n$  matrix  $\Gamma$  is the covariance matrix of some set of random variables  $Y_1, \dots, Y_n$  if and only if

- 1)  $\Gamma$  is symmetric nonnegative definite, equivalently,
- 2) there is a matrix  $A$  such that

$$\Gamma = A^t A.$$

*Proof.* One way is easy. Suppose  $\Gamma_{ij} = EY_i Y_j$ ; then obviously  $\Gamma$  is symmetric, and  $\sum \alpha_i \Gamma_{ij} \alpha_j = E(\sum \alpha_i Y_i)^2 \geq 0$ . For the converse, we start with the well-known result (see [63], for example) that for  $\Gamma$  symmetric and nonnegative definite, there is an orthogonal matrix  $O$  such that

$$O\Gamma O^t = D,$$

where  $D$  is diagonal with nonnegative elements. Then taking  $B$  to be diagonal with diagonal elements the square roots of the corresponding elements  $D$  gives  $B = D^t D$  so that

$$\Gamma = (O^t D^t)(DO).$$

Take  $A = DO$ , then  $\Gamma = A^t A$ . Now take  $\mathbf{X}$  with components independent  $\mathcal{N}(0, 1)$  variables,  $\mathbf{Y} = \mathbf{XA}$  to get the result that the covariance matrix of  $\mathbf{Y}$  is  $\Gamma$ .

If  $\mathbf{Y}$  has joint normal distribution,  $\mathbf{Y} = \mathbf{XA}$ , then

$$(\mathbf{u}, \mathbf{Y}) = \sum_j u_j Y_j = \sum_{j,k} u_j \mathbf{X}_k a_{kj}.$$

Since the  $\mathbf{X}_k$  are independent  $\mathcal{N}(0, 1)$ ,

$$\begin{aligned} Ee^{i(\mathbf{u}, \mathbf{Y})} &= E \exp \left[ i \sum_k \mathbf{X}_k \left( \sum_j u_j a_{kj} \right) \right] \\ &= \exp \left[ -\frac{1}{2} \sum_k \left( \sum_j u_j a_{kj} \right)^2 \right] = \exp \left[ -\frac{1}{2} \mathbf{u} A^t A \mathbf{u}^t \right]. \end{aligned}$$

Hence the  $\mathbf{Y}$  vector has a joint normal distribution in the sense of 11.9. Furthermore, if  $\mathbf{Y}$  has the characteristic function  $e^{-(1/2)\mathbf{u}\Gamma\mathbf{u}^t}$ , by finding  $\mathbf{A}$  such that  $\Gamma = \mathbf{A}^t\mathbf{A}$ , taking  $\tilde{\mathbf{X}}$  to have independent  $\mathcal{N}(0, 1)$  components and  $\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\mathbf{A}$ , then we get  $\tilde{\mathbf{Y}} \stackrel{d}{=} \mathbf{Y}$ .

We can do better. Suppose  $\mathbf{Y}$  has characteristic function  $e^{-(1/2)\mathbf{u}\Gamma\mathbf{u}^t}$ . Take  $\mathbf{O}$  to be an orthogonal matrix such that  $\mathbf{O}\Gamma\mathbf{O}^t = \mathbf{D}$  is diagonal. Consider the vector  $\mathbf{Z} = \mathbf{Y}\mathbf{O}$ .

$$E(\mathbf{Z}_i\mathbf{Z}_j) = E\left(\sum_{k,m} Y_k O_{ki} Y_m O_{mj}\right) = \sum_{k,m} O_{ki} \Gamma_{km} O_{mj}.$$

So

$$E(\mathbf{Z}_i\mathbf{Z}_j) = \begin{cases} 0, & i \neq j, \\ d_{jj}, & i = j. \end{cases}$$

Thus, the characteristic function of  $\mathbf{Z}$  splits into a product, the  $\mathbf{Z}_1, \mathbf{Z}_2, \dots$  are independent by 8.33, and  $\mathbf{Z}_j$  is  $\mathcal{N}(0, d_{jj})$ . Define  $\mathbf{X}_j \equiv 0$  if  $d_{jj} = 0$ , otherwise  $\mathbf{X}_j = \mathbf{Z}_j/d_{jj}$ . Then the nonzero  $\mathbf{X}_j$  are independent  $\mathcal{N}(0, 1)$  variables, and there is a matrix  $\mathbf{A}$  such that  $\mathbf{Y} = \mathbf{X}\mathbf{A}$ .

A fresh memory of linear algebra will show that all that has been done here is to get an orthonormal basis for the functions  $Y_1, \dots, Y_n$ . This could be done for any set of random variables  $Y_1, \dots, Y_n$ , getting random variables  $X_1, \dots, X_k$  such that  $EX_j^2 = 1$ ,  $EX_i X_j = 0$ ,  $i \neq j$ . But the variables  $Y_1, \dots, Y_n$ , having a joint normal distribution and zero means, have their distribution completely determined by  $\Gamma$  with the pleasant and unusual property that  $EY_i Y_j = 0$  implies  $Y_i$  and  $Y_j$  are independent. Furthermore, if  $I_1 = (i_1, \dots, i_k)$ ,  $I_2 = (j_1, \dots, j_m)$  are disjoint subsets of  $(1, \dots, n)$  and  $\Gamma_{ij} = 0$ ,  $i \in I_1, j \in I_2$ , then it follows that  $(Y_{i_1}, \dots, Y_{i_k})$  and  $(Y_{j_1}, \dots, Y_{j_m})$  are independent vectors.

We make an obvious extension to nonzero means by

**Definition 11.14.** Say that  $\mathbf{Y} = (Y_1, \dots, Y_n)$ , with  $\mathbf{m} = (EY_1, \dots, EY_n)$ , has a joint normal distribution  $\mathcal{N}(\mathbf{m}, \Gamma)$  if  $(Y_1 - EY_1, \dots, Y_n - EY_n)$  has the joint normal distribution  $\mathcal{N}(\mathbf{0}, \Gamma)$ .

Also,

**Definition 11.15.** If  $\mathbf{Y} = (Y_1, \dots, Y_n)$  has the distribution  $\mathcal{N}(\mathbf{0}, \Gamma)$ , the distribution is said to be nondegenerate if the  $Y_1, \dots, Y_n$  are linearly independent, or equivalently, if the rank of  $\Gamma$  is  $n$ .

## Problems

5. Show that if  $Y_1, \dots, Y_n$  are  $\mathcal{N}(0, \Gamma)$ , the distribution is nondegenerate if and only if there are  $n$  independent random variables  $X_1, \dots, X_n$ , each  $\mathcal{N}(0, 1)$ , and an  $n \times n$  matrix  $\mathbf{A}$  such that  $\det(\mathbf{A}) \neq 0$ , and  $\mathbf{Y} = \mathbf{X}\mathbf{A}$ .

6. Let  $Y_1, \dots, Y_n$  have a joint normal nondegenerate distribution. Show that their distribution function has a density  $f(\mathbf{y})$  given by

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Gamma)}} \exp \left[ -\frac{1}{2} \mathbf{y} \Gamma^{-1} \mathbf{y}' \right].$$

7. Show that if the random variables  $Y_1, \dots, Y_n$  have the characteristic function  $\exp \left[ -\frac{1}{2} \mathbf{u} H \mathbf{u}' \right]$  for some  $n \times n$  matrix  $H$ , then  $H$  is the covariance matrix of  $\mathbf{Y}$  and  $EY_j = 0, j = 1, \dots, n$ .

## 5. STATIONARY GAUSSIAN PROCESS

**Definition 11.16.** A double ended process  $\dots, X_{-1}, X_0, X_1, \dots$  is called Gaussian if every finite subset of variables has a joint normal distribution.

Of course, this assumes that  $E|X_n| < \infty$  for all  $n$ . When is a Gaussian zero-mean process  $\mathbf{X}$  stationary? Take  $\Gamma(m, n) = EX_n X_m$ . Since the distribution of the process is determined by  $\Gamma(m, n)$ , the condition should be on  $\Gamma(m, n)$ .

**Proposition 11.17.**  $\mathbf{X}$  is stationary if and only if

$$\Gamma(m, n) = \Gamma(m - n, 0), \quad \text{all } m, n.$$

*Proof.* If  $\mathbf{X}$  is stationary then  $EX_m X_n = EX_{m-n} X_0 = \Gamma(m - n, 0)$ . Conversely, if true, then the characteristic function of  $X_1, \dots, X_n$  is

$$\exp \left( \frac{1}{2} \sum_{k,j=1}^n u_j u_k \Gamma(k - j, 0) \right).$$

But this is exactly the characteristic function of  $X_{1+m}, \dots, X_{n+m}$ .

Use the notation (loose),

$$\Gamma(n) = \Gamma(n, 0)$$

and call  $\Gamma(n)$  the covariance function. Call a function  $M(n)$  on the integers nonnegative definite if for  $I$  any finite subset of the integers, and  $\alpha_j, j \in I$ , any real numbers,

$$\sum_{k,j \in I} \alpha_k \alpha_j M(k - j) \geq 0.$$

Clearly a covariance function is nonnegative definite. Just as in the finite case, given any symmetric nonnegative definite function  $H(n)$  on the integers, we can construct a stationary zero-mean Gaussian process such that  $EX_m X_n = H(m - n)$ .

How can we describe the general stationary Gaussian process? To do this neatly, generalize a bit. Let  $\dots, X_{-1}, X_0, X_1, \dots$  be a process of complex-valued functions  $X_j = U_j + iV_j$ , where  $U_j, V_j$  are random variables, and  $EU_j = EV_j = 0$ , all  $j$ . Call it a complex Gaussian process if any finite

subset of the  $\{U_n, V_n\}$  have a joint normal distribution, stationary if  $EX_m \bar{X}_n = \Gamma(m - n)$ . The covariance function of a complex Gaussian stationary process is Hermitian,  $\Gamma(-n) = \bar{\Gamma}(n)$  and nonnegative definite in the sense that for a subset  $I$  of the integers, and  $\alpha_j$  complex numbers,

$$\sum_{k, j \in I} \alpha_k \bar{\alpha}_j \Gamma(k - j) = E \left( \left| \sum_{k \in I} \alpha_k X_k \right|^2 \right) \geq 0.$$

Consider a process that is a superposition of periodic functions with random amplitudes that are independent and normal. More precisely, let  $\lambda_1, \dots, \lambda_k$  be real numbers (called frequencies), and define

$$(11.18) \quad X_n = \sum_{j=1}^k e^{i\lambda_j n} Z_j,$$

where  $Z_1, \dots, Z_k$  are independent  $\mathcal{N}(0, \sigma_j^2)$  variables. The  $\{X_n\}$  process is a complex Gaussian process. Further,

$$\begin{aligned} EX_n \bar{X}_m &= \sum_j \sum_m e^{i\lambda_j n} e^{-i\lambda_j m} E Z_j Z_l \\ &= \sum_j \sigma_j^2 e^{i\lambda_j (n-m)}, \end{aligned}$$

so the process is stationary. The functions  $e^{i\lambda_j n}$  are the periodic components with frequency  $\lambda_j$ , and we may as well take  $\lambda_j \in [-\pi, +\pi)$ . The formula (11.18) can be thought of as representing the  $\{X_n\}$  process by a sum over frequency space  $\lambda$ ,  $\lambda \in [-\pi, +\pi)$ . The main structural theorem for complex normal stationary processes is that every such process can be represented as an integral over frequency space, where the amplitudes of the various frequency components are, in a generalized sense, independent and normally distributed.

## 6. SPECTRAL REPRESENTATION OF STATIONARY GAUSSIAN PROCESSES

The main tool in the representation theorem is a representation result for covariance functions.

**Herglotz Lemma 11.19** [70].  $\Gamma(n)$  is a Hermitian nonnegative definite function on the integers if and only if there is a finite measure  $F(B)$  on  $\mathcal{B}_1([-\pi, +\pi))$  such that

$$\Gamma(n) = \int e^{in\lambda} F(d\lambda).$$

*Proof.* One direction is quick. If

$$\Gamma(n) = \int e^{in\lambda} F(d\lambda),$$

then  $\Gamma(n)$  is Hermitian, and

$$\sum_{k, j \in I} \alpha_k \bar{\alpha}_j \Gamma(k - j) = \int |\sum \alpha_k e^{ik\lambda}|^2 F(d\lambda) \geq 0.$$

To go the other way, following Loève [108, p. 207], define

$$\begin{aligned} f_n(\lambda) &= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \left(1 - \frac{|k|}{n}\right) \Gamma(k) e^{-ik\lambda} \\ &= \frac{1}{2\pi n} \sum_{j=1}^n \sum_{m=1}^n \Gamma(j - m) e^{-ij\lambda} e^{im\lambda} \geq 0. \end{aligned}$$

Multiply both sides by  $e^{il}$  for  $-n+1 \leq l \leq n-1$ , and integrate over  $[-\pi, +\pi]$  to get

$$\begin{aligned} \left(1 - \frac{|l|}{n}\right) \Gamma(l) &= \int e^{il\lambda} f_n(\lambda) d\lambda, \\ \Gamma(0) &= \int f_n(\lambda) d\lambda. \end{aligned}$$

Define  $F_n(d\lambda)$  as the measure on  $[-\pi, +\pi]$  with density  $f_n(\lambda)$ , and take  $n'$  a subsequence such that  $F_{n'} \xrightarrow{\mathcal{D}} F$ . Then

$$\begin{aligned} \Gamma(l) &= \lim \int e^{il\lambda} F_{n'}(d\lambda) \\ &= \int e^{il\lambda} F(d\lambda). \end{aligned}$$

Now take the mass on the point  $\{\pi\}$  and put it on  $\{-\pi\}$  to complete the proof.

Note that the functions  $\{e^{in\lambda}\}$  are separating on  $[-\pi, \pi)$ , hence  $F(d\lambda)$  is uniquely determined by  $\Gamma(n)$ . For  $\Gamma(n)$  the covariance function of a complex Gaussian stationary process  $\mathbf{X}$ ,  $F(d\lambda)$  is called the *spectral distribution function* of  $\mathbf{X}$ . To understand the representation of  $\mathbf{X}$ , an integral with respect to a random measure has to be defined.

**Definition 11.20.** Let  $\{Z(\lambda)\}$  be a noncountable family of complex-valued random variables on  $(\Omega, \mathcal{F}, P)$  indexed by  $\lambda \in [-\pi, +\pi)$ . For  $I$  an interval  $[\lambda_1, \lambda_2)$ , define  $Z(I) = Z(\lambda_2) - Z(\lambda_1)$ . If the Riemann sums  $\sum f(\lambda_k) Z(I_k)$ ,  $I_1, \dots, I_n$  a disjoint partition of  $[-\pi, +\pi)$  into intervals left-closed, right-open,  $\lambda_k \in I_k$ , converge in the second mean to the same random variable for any sequence of partitions such that  $\max_k |I_k| \rightarrow 0$ , denote this limit random variable by

$$\int f(\lambda) Z(d\lambda).$$

Now we can state



**Theorem 11.21.** Let  $X$  be a complex Gaussian stationary process on  $(\Omega, \mathcal{F}, P)$  with spectral distribution function  $F(d\lambda)$ . Then there exists a family  $\{Z(\lambda)\}$  of complex-valued random variables on  $(\Omega, \mathcal{F}, P)$  indexed by  $\lambda \in [-\pi, +\pi)$  such that

- i) for any  $\lambda_1, \dots, \lambda_m$ ,  $Z(\lambda_1), \dots, Z(\lambda_m)$  have a joint normal distribution,
- ii) for  $I_1, I_2$ , disjoint,  $EZ(I_1)\overline{Z(I_2)} = 0$ ,
- iii)  $E|Z(I)|^2 = F(I)$ , for all intervals  $I$ ,
- iv)  $X_n = \int e^{in\lambda} Z(d\lambda)$ , a.s. all  $n$ .

*Proof.* The most elegant way to prove this is to use some elementary Hilbert space arguments. Consider the space  $\mathcal{L}(X)$  consisting of all finite linear combinations

$$\sum \alpha_k X_k,$$

where the  $\alpha_k$  are complex numbers. Consider the class of all random variables  $Y$  such that there exists a sequence  $Y_n \in \mathcal{L}$  with  $Y_n \xrightarrow{2} Y$ . On this class define an inner product  $(Y_1, Y_2)$  by  $EY_1\overline{Y_2}$ . Call random variables equivalent if they are a.s. equal. Then it is not difficult to check that the set of equivalence classes of random variables forms a complete Hilbert space  $L_2(X)$  under the inner product  $(Y_1, Y_2)$ . Let  $L_2(F)$  be the Hilbert space of all complex-valued  $\mathcal{B}_1([-\pi, +\pi))$  measurable functions  $f(\lambda)$  such that  $\int |f(\lambda)|^2 F(d\lambda) < \infty$  under the inner product  $(f, g) = \int fg \, dF$  (take equivalence classes again).

To the element  $X_n \in L_2(X)$ , correspond the function  $e^{in\lambda} \in L_2(F)$ . Extend this correspondence linearly,

$$\sum \alpha_k X_k \leftrightarrow \sum \alpha_k e^{ik\lambda}.$$

Let  $\mathcal{L}(F)$  be the class of all finite linear combinations  $\sum \alpha_k e^{ik\lambda}$ . Then

$$Y_1, Y_2 \in \mathcal{L}(X), f, g \in \mathcal{L}(F),$$

and  $Y_1 \leftrightarrow f, Y_2 \leftrightarrow g$  implies  $\alpha Y_1 + \beta Y_2 \leftrightarrow \alpha f + \beta g$ , and

$$\begin{aligned} (Y_1, Y_2) &= E(\sum \alpha_j^{(1)} X_j) \overline{(\sum \alpha_k^{(2)} X_k)} = \sum \alpha_j^{(1)} \overline{\alpha_k^{(2)}} \Gamma(j - k) \\ &= \int \sum \alpha_j^{(1)} \overline{\alpha_k^{(2)}} e^{i(j-k)\lambda} F(d\lambda) = \int f(\lambda) \overline{g(\lambda)} F(d\lambda) = (f, g). \end{aligned}$$

If  $Y_n \in \mathcal{L}(X)$  and  $Y_n \xrightarrow{2} Y$ , then  $Y_n$  is Cauchy-convergent in the second mean; consequently so is the sequence  $f_n \leftrightarrow Y_n$ . Hence there is an  $f \in L_2(F)$  such that  $f_n \xrightarrow{2} f$ . Define  $Y \leftrightarrow f$ ; this can be checked to give a one-to-one correspondence between  $L_2(F)$  and  $L_2(X)$ , which is linear and preserves inner products.

The function  $\chi_{[-\pi, \xi]}(\lambda)$  is in  $L_2(F)$ ; let  $Z(\xi)$  be the corresponding element in  $L_2(X)$ . Now to check that the family  $\{Z(\xi)\}$  has the properties asserted. Begin with (i). If real random variables  $Y_k^{(n)} \xrightarrow{2} Y_k$ ,  $k = 1, \dots, m$ , and  $(Y_1^{(n)}, \dots, Y_m^{(n)})$  has a joint normal distribution for each  $n$ , then  $(Y_1, \dots, Y_m)$  is joint normal. Because each element  $\Gamma_{k,j}^{(n)} = EY_k^{(n)}Y_j^{(n)}$  of the covariance matrix converges to  $\Gamma_{k,j} = EY_kY_j$ ; hence the characteristic function of  $Y_n$  converges to  $e^{-(1/2)u^t \Gamma u}$ , and this must be the characteristic function of  $Y$ . Conclude that if  $Y_1, \dots, Y_m$  are in  $L_2(X)$ , then their real and imaginary components have joint normal distributions. Thus for  $\xi_1, \dots, \xi_m$  in  $[-\pi, \pi)$ , the real and imaginary components of  $Z(\xi_1), \dots, Z(\xi_m)$  have a joint normal distribution. For any interval  $I = [\xi_1, \xi_2)$ ,  $Z(I) \leftrightarrow \chi_I(\lambda)$ . Hence for  $I_1, I_2$  disjoint,

$$EZ(I_1)\overline{Z(I_2)} = \int \chi_{I_1}(\lambda)\overline{\chi_{I_2}(\lambda)}F(d\lambda) = 0.$$

Also,

$$E|Z(I)|^2 = \int \chi_I^2(\lambda)F(d\lambda) = \int \chi_I(\lambda)F(d\lambda) = F(I).$$

Lastly, take  $f(\lambda)$  to be a uniformly continuous function on  $[-\pi, \pi)$ . For a partition of  $[-\pi, \pi)$  into disjoint intervals  $I_1, \dots, I_k$  left-closed, right-open, and  $\lambda_k \in I_k$ ,

$$\Sigma f(\lambda_k)Z(I_k) \leftrightarrow \Sigma f(\lambda_k)\chi_{I_k}(\lambda).$$

The function on the right equals  $f(\lambda_k)$  on the interval  $I_k$ , and converges uniformly to  $f(\lambda)$  as  $\max_k |I_k| \rightarrow 0$ . So, in particular, it converges in the second mean to  $f(\lambda)$ . If  $Y \leftrightarrow f(\lambda)$ , then  $\Sigma f(\lambda_k)Z(I_k) \xrightarrow{2} Y$ . From the definition 11.20,

$$Y = \int f(\lambda)Z(d\lambda).$$

For  $f(\lambda) = e^{in\lambda}$ , the corresponding element is  $X_n$ , thus

$$X_n = \int e^{in\lambda}Z(d\lambda) \quad \text{a.s.} \quad \text{Q.E.D.}$$

**Proposition 11.22.** *If  $\{X_n\}$  is a real stationary Gaussian process, then the family  $\{Z(\lambda)\}$ ,  $\lambda \in [-\pi, +\pi)$  has the additional properties: If  $\{Z_1(\lambda)\}$ ,  $\{Z_2(\lambda)\}$  are the real and imaginary parts of the  $\{Z(\lambda)\}$  process*

$$Z(\lambda) = Z_1(\lambda) + iZ_2(\lambda),$$

then

i) For any two intervals  $I, J$ ,

$$EZ_1(I)Z_2(J) = 0.$$

ii) For any two disjoint intervals  $I, J$ ,

$$EZ_1(I)Z_1(J) = 0, \quad EZ_2(I)Z_2(J) = 0.$$

*Proof.* Write any intervals  $I, J$  as the union of the common part  $I \cap J$  and nonoverlapping parts, and apply 11.21(ii) and (iii) to conclude that the imaginary part of  $EZ(I)\overline{Z(J)}$  is zero. Therefore,

$$(11.23) \quad EZ_1(I)Z_2(J) + EZ_2(I)Z_1(J) = 0.$$

Inspect the correspondence set up in the proof of 11.21 and notice that if  $Y = \sum \alpha_k X_k$  and  $Y \leftrightarrow f(\lambda)$ , then  $\bar{Y} = \sum \bar{\alpha}_k X_k$  corresponds to  $\overline{f(-\lambda)}$ . This extends to all  $L_2(\mathbf{X})$  and  $L_2(F)$ . Hence, since  $\chi_I(\lambda) \leftrightarrow Z(I)$ , then  $\chi_{-I}(\lambda) \leftrightarrow \overline{Z(I)}$ ; so

$$Z(-I) = \overline{Z(I)} \quad \text{a.s.,}$$

where  $-I = \{\lambda; -\lambda \in I\}$ . From this,

$$Z_1(-I) = Z_1(I), \quad Z_2(-I) = -Z_2(I).$$

Thus, if we change  $I$  to  $-I$  in (11.23), the first term remains the same, and the second term changes sign; hence both terms are zero. For  $I, J$  disjoint,  $EZ(I)\overline{Z(J)} = 0$  implies  $EZ_1(I)Z_1(J) = EZ_2(I)Z_2(J)$ . We can use the sign change again as above to prove that both sides are individually zero. Q.E.D.

For a real stationary Gaussian process with zero means, we can deduce from 11.22(i) that the processes  $\{Z_1(\lambda)\}, \{Z_2(\lambda)\}$  are independent in the sense that all finite subsets

$$\{Z_1(I_j)\}, \quad j = 1, \dots, n, \quad \{Z_2(I'_k)\}, \quad k = 1, \dots, m,$$

are independent. From 11.22 (ii), we deduce that for  $I_1, \dots, I_n$  disjoint intervals, the random variables  $Z_1(I_1), \dots, Z_1(I_n)$  are independent. Similarly for  $Z_2(I_1), \dots, Z_2(I_n)$ .

## 7. OTHER PROBLEMS

The fact that  $F(d\lambda)$  completely determines the distribution of a stationary Gaussian process with zero means leads to some compact results. For example, the process  $\mathbf{X}$  is ergodic if and only if  $F(d\lambda)$  assigns no mass to any one-point sets [111].

The correspondence between  $L_2(\mathbf{X})$  and  $L_2(F)$  was exploited by Kolmogorov [96, 97] and independently by Wiener [143] in a fascinating piece of analysis that leads to the solution of the prediction problem. The starting point is this: the best predictor in a mean-square sense of  $X_1$  based on  $X_0, X_{-1}, \dots$  is  $E(X_1 | X_0, X_{-1}, \dots)$ . But for a Gaussian process, there are constants  $\alpha_k^{(n)}$  such that

$$E(X_1 | X_0, \dots, X_{-n}) = \sum_0^n \alpha_k^{(n)} X_{-k}.$$

Because by taking the  $\alpha_k^{(n)}$  such that

$$E\left[\left(X_1 - \sum_0^n \alpha_k^{(n)} X_{-k}\right) X_j\right] = 0, \quad j = 0, \dots, -n,$$

or

$$\Gamma_{1,j} = \sum_0^n \alpha_k^{(n)} \Gamma_{-k,j}, \quad j = 0, \dots, -n,$$

then  $X_1 - \sum_1^n \alpha_k^{(n)} X_{-k}$  is independent of  $(X_0, X_{-1}, \dots, X_{-n})$ , so that

$$E\left(X_1 - \sum_0^n \alpha_k^{(n)} X_{-k} \mid X_0, X_{-1}, \dots, X_{-n}\right) = 0.$$

From the Martingale theorems, since  $EX_1^2 < \infty$ , it is easy to deduce that

$$E(X_1 \mid X_0, \dots, X_{-n}) \xrightarrow{2} E(X_1 \mid X_0, \dots).$$

Hence the best predictor is in the space  $L_2(X_0, X_{-1}, \dots)$  generated by all linear combinations of  $X_0, X_{-1}, \dots$ . By the isomorphism this translates into the problem: Let  $L_2^-(F)$  be the space generated by all linear combinations of  $e^{ik\lambda}$ ,  $k = 0, -1, -2, \dots$ . Find the element  $f(\lambda) \in L_2^-(F)$  which minimizes

$$\int |e^{i\lambda} - f(\lambda)|^2 F(d\lambda).$$

In a similar way, many problems concerning Gaussian processes translate over into interesting and sometimes well-known problems in functions of a real variable, in particular, usually in the area of approximation theory.

## NOTES

Note that the representation theorem 11.21 for Gaussian processes depends only on the fact that  $EX_n \bar{X}_m$  depends on the difference  $n - m$ . Define  $\{X_n\}$ ,  $n = 0, \pm 1, \dots$  to be a complex process *stationary in the second order* if  $\Gamma(n, m) = EX_n \bar{X}_m$  is a function of  $n - m$ . The only difference in the conclusion of 11.21 is that (i) is deleted. This representation theorem was proved by Cramér [21, 1942], and independently by Loève [107]. Since the work on the prediction problem in 1941–42 by Kolmogorov [96] and [97], and independently by Wiener [143], there has been a torrent of publications on second-order stationary processes and a sizeable amount on Gaussian processes. For a complete and rigorous treatment of these matters, refer to Doob's book [39]. For a treatment which is simpler and places more stress on applications, see Yaglom [144].

## CHAPTER 12

# STOCHASTIC PROCESSES AND BROWNIAN MOTION

### 1. INTRODUCTION

The natural generalization of a sequence of random variables  $\{X_n\}$  is a collection of random variables  $\{X_t\}$  indexed by a parameter  $t$  in some interval  $I$ . Such an object we will call a stochastic process.

**Definition 12.1.** A stochastic process or continuous parameter process is a collection  $\{X_t(\omega)\}$  of random variables on  $(\Omega, \mathcal{F}, P)$  where  $t$  ranges over an interval  $I \subset \mathbb{R}^{(1)}$ . Whenever convenient the notation  $\{X(t, \omega)\}$  or simply  $\{X(t)\}$  will be used.

For fixed  $\omega$ , what is produced by observing the values of  $X(t, \omega)$  is a function  $x(t)$  on  $I$ .

The most famous stochastic process and the most central in probability theory is Brownian motion. This comes up like so: let  $X(t)$  denote one position coordinate of a microscopic particle undergoing molecular bombardments in a glass of water. Make the three assumptions given below.

#### Assumptions 12.2

- 1) *Independence:*  $X(t + \Delta t) - X(t)$  is independent of  $\{X(\tau)\}, \tau \leq t$ .
- 2) *Stationarity:* The distribution of  $X(t + \Delta t) - X(t)$  does not depend on  $t$ .
- 3) *Continuity:* 
$$\lim_{\Delta t \downarrow 0} \frac{P(|X(t + \Delta t) - X(t)| \geq \delta)}{\Delta t} = 0, \text{ for all } \delta > 0.$$

This is the sense of the assumptions: (1) means that the change in position during time  $[t, t + \Delta t]$  is independent of anything that has happened up to the time  $t$ . This is obviously only a rough approximation. Physically, what is much more correct is that the momentum imparted to the particle due to molecular bombardments during  $[t, t + \Delta t]$  is independent of what has happened up to time  $t$ . This assumption makes sense only if the

displacement of the particle due to its initial velocity at the beginning of the interval  $[t, t + \Delta t]$  is small compared to the displacements it suffers as a result of molecular momentum exchange over  $[t, t + \Delta t]$ . From a model point of view this is the worst assumption of the three. Accept it for now; later we derive the so-called exact model for the motion in which (1) will be replaced. The second assumption is quite reasonable: It simply requires homogeneity in time; that the distribution of change over any time interval depend only on the length of the time interval, and not on the location of the origin of time. This corresponds to a model in which the medium is considered to be infinite in extent.

The third assumption is interesting. We want all the sample functions of our motion to be continuous. A model in which the particle took instantaneous jumps would be a bit shocking. Split the interval  $[0, 1]$  into  $n$  parts,  $\Delta t = 1/n$ . If the motion is continuous, then

$$h(\Delta t) = \sup_{1 \leq k \leq n} |X(k \Delta t) - X((k-1) \Delta t)|$$

must converge to zero as  $\Delta t \rightarrow 0$ . At a minimum, for any  $\delta > 0$ ,

$$\lim_{\Delta t \downarrow 0} P(h(\Delta t) \geq \delta) = 0.$$

By (1) the variables  $Y_k = |X(k \Delta t) - X((k-1) \Delta t)|$  are independent; by (2), they all have the same distribution. Thus

$$\begin{aligned} (12.3) \quad P(h(\Delta t) \geq \delta) &= 1 - P\left(\sup_{1 \leq k \leq n} Y_k < \delta\right) = 1 - [P(Y_1 < \delta)]^n \\ &= 1 - [1 - P(Y_1 \geq \delta)]^n \simeq 1 - e^{-nP(Y_1 \geq \delta)}, \end{aligned}$$

so that  $P(h(\Delta t) \geq \delta) \rightarrow 0$  if and only if  $nP(Y_1 \geq \delta) \rightarrow 0$ . This last is exactly

$$\lim_{n \rightarrow \infty} \frac{P(|X(t + \Delta t) - X(t)| \geq \delta)}{\Delta t} = 0.$$

Make the further assumption that  $X(0) \equiv 0$ . This is not a restriction, but can be done by considering the process  $X(t) - X(0)$ ,  $t \geq 0$ , which again satisfies (1), (2), and (3). Then

**Proposition 12.4.** *For any process  $\{X(t)\}$ ,  $t \geq 0$  satisfying 12.2. (1), (2), and (3) with  $X(0) \equiv 0$ ,  $X(t)$  has a normal distribution with  $EX(t) = \mu t$ ,  $\sigma^2(X(t)) = \sigma^2 t$ .*

*Proof.* For any  $t$ , let  $\Delta t = t/n$ ,  $Y_k = X(k \Delta t) - X((k-1) \Delta t)$ . Then  $X(t) = Y_1 + \cdots + Y_n$ , where the  $Y_1, \dots, Y_n$  are independent and identically distributed. Therefore  $X(t)$  has an infinitely divisible law. Utilize the proof in (12.3) to show that  $M_n = \max_{1 \leq k \leq n} |Y_k|$  converges in probability to

zero. By 9.6,  $X(t)$  has a normal distribution. Let  $\varphi_1(t) = EX(t)$ . Then

$$\begin{aligned}\varphi_1(t + \tau) &= EX(t + \tau) \\ &= E(X(t + \tau) - X(t)) + EX(t) \\ &= \varphi_1(t) + \varphi_1(\tau).\end{aligned}$$

Let

$$\varphi_2(t) = \sigma^2(X(t)),$$

so that

$$\begin{aligned}\varphi_2(t + \tau) &= E(X(t + \tau) - \varphi_1(t + \tau))^2 \\ &= E(X(t + \tau) - X(t) - \varphi_1(\tau) + X(t) - \varphi_1(t))^2 \\ &= \varphi_2(t) + \varphi_2(\tau).\end{aligned}$$

The fact is now that  $\varphi_1(t)$  and  $\varphi_2(t)$  are continuous. This follows from  $X(t + \tau) \xrightarrow{D} X(t)$ , as  $\tau \rightarrow 0$ , which for normal variables implies

$$EX(t + \tau) \rightarrow EX(t), \quad \sigma^2(X(t + \tau)) \rightarrow \sigma^2(X(t)).$$

It is easy to show that any continuous solutions of the equation  $\varphi(t + \tau) = \varphi(t) + \varphi(\tau)$  are linear.

Use the above heuristics to back into a definition of Brownian motion.

**Definition 12.5.** *Brownian motion is a stochastic process on  $[0, \infty)$  such that  $X(0) \equiv 0$  and the joint distribution of*

$$X(t_n), \dots, X(t_0), \quad t_n > t_{n-1} > \dots > t_0 \geq 0,$$

*is specified by the requirement that  $X(t_k) - X(t_{k-1})$ ,  $k = 1, \dots, n$  be independent, normally distributed random variables with*

$$\begin{aligned}E(X(t_k) - X(t_{k-1})) &= (t_k - t_{k-1})\mu, \\ \sigma^2(X(t_k) - X(t_{k-1})) &= (t_k - t_{k-1})\sigma^2.\end{aligned}$$

This can be said another way. The random variables  $X(t_n), \dots, X(t_0)$  have a joint normal distribution with  $EX(t_k) = \mu t_k$  and

$$\begin{aligned}\Gamma(t_j, t_k) &= E(X(t_j) - \mu t_j)(X(t_k) - \mu t_k) \\ &= E[(X(t_j) - X(t_k) - \mu(t_j - t_k) + X(t_k) - \mu t_k)(X(t_k) - \mu t_k)] \\ &= \sigma^2 t_k, \quad \text{for } t_j > t_k,\end{aligned}$$

so that  $\Gamma(t_j, t_k) = \sigma^2(\min(t_j, t_k))$ .

**Problem 1.** Show that Brownian motion, as defined by 12.5, satisfies 12.2 (1), (2), and (3).

## 2. BROWNIAN MOTION AS THE LIMIT OF RANDOM WALKS

There are other ways of looking at Brownian motion. Consider a particle that moves to the right or left a distance  $\Delta x$  with probability  $\frac{1}{2}$ . It does this each  $\Delta t$  time unit. Let  $Y_1, \dots$  be independent, and equal  $\pm \Delta x$  with probability  $\frac{1}{2}$  each. The particle at time  $t$  has made  $[t/\Delta t]$  jumps ( $[z]$  indicates greatest integer  $\leq z$ ). Thus the position of the particle is given by

$$D(t) = Y_1 + \cdots + Y_{[t/\Delta t]}.$$

The idea is that if  $\Delta x, \Delta t \rightarrow 0$  in the right way, then  $D(t)$  will approach Brownian motion in some way. To figure out how to let  $\Delta x, \Delta t \rightarrow 0$ , note that  $ED^2(t) \simeq (\Delta x)^2 t / \Delta t$ . To keep this finite and nonzero,  $\Delta x$  has to be of the order of magnitude of  $\sqrt{\Delta t}$ . For simplicity, take  $\Delta x = \sqrt{\Delta t}$ . Take  $\Delta t = 1/n$ , then the  $Y_1, \dots$  equal  $\pm 1/\sqrt{n}$ . Thus the  $D(t)$  process has the same distribution as

$$X^{(n)}(t) = \frac{Z_1 + \cdots + Z_{[nt]}}{\sqrt{n}},$$

where  $Z_1, \dots$  are  $\pm 1$  with probability  $\frac{1}{2}$ .

Note that

$$X^{(n)}(t) = \sqrt{t} \frac{Z_1 + \cdots + Z_{[nt]}}{\sqrt{nt}}$$

and apply the central limit theorem to conclude  $X^{(n)}(t) \xrightarrow{\mathcal{D}} \mathcal{N}(0, t)$ . In addition, it is no more difficult to show that all the joint distributions of  $X^{(n)}(t)$  converge to those of Brownian motion. Therefore, Brownian motion appears as the limit of processes consisting of consecutive sums of independent, identically distributed random variables, and its study is an extension of the study of the properties of such sequences.

What has been done in 12.5 is to specify all the finite-dimensional distribution functions of the process. There is now the question again: Is there a process  $\{X(t)\}$ ,  $t \in [0, \infty)$  on  $(\Omega, \mathcal{F}, P)$  with these finite-dimensional distributions? This diverts us into some foundational work.

## 3. DEFINITIONS AND EXISTENCE

Consider a stochastic process  $\{X(t)\}$ ,  $t \in I$  on  $(\Omega, \mathcal{F}, P)$ . For fixed  $\omega$ ,  $X(t, \omega)$  is a real-valued function on  $I$ . Hence denote by  $R^I$  the class of all real-valued functions  $x(t)$  on  $I$ , and by  $X(\omega)$  the vector variable  $\{X(t, \omega)\}$  taking values in  $R^I$ .

**Definition 12.6.**  $\mathcal{F}(X(s), s \in J)$ ,  $J \subset I$ , is the smallest  $\sigma$ -field  $\mathcal{F}$  such that all  $X(s)$ ,  $s \in J$ , are  $\mathcal{F}$ -measurable.

**Definition 12.7.** A finite-dimensional rectangle in  $R^I$  is any set of the form

$$S = \{x(\cdot) \in R^I; x(t_1) \in I_1, \dots, x(t_n) \in I_n\},$$



where  $I_1, \dots, I_n$  are intervals. Let  $\mathcal{B}_I$  be the smallest  $\sigma$ -field of subsets of  $R^I$  containing all finite dimensional rectangles.

For the understanding of what is going on here it is important to characterize  $\mathcal{B}_I$ . Say that a set  $B \in \mathcal{B}_I$  has a countable base  $T = \{t_j\}$  if it is of the form

$$B = \{x(\cdot); (x(t_1), x(t_2), \dots) \in D\}, \quad D \in \mathcal{B}_\infty.$$

This means that  $B$  is a set depending only on the coordinates  $x(t_1), x(t_2), \dots$

**Proposition 12.8.** *The class  $\mathcal{C}$  of all sets with a countable base forms a  $\sigma$ -field, hence  $\mathcal{C} = \mathcal{B}_I$ .*

*Proof.* Let  $B_1, B_2, \dots \in \mathcal{C}$ , and take  $T_k$  as the base for  $B_k$ , then  $T = \bigcup_k T_k$  is a base for all  $B_1, B_2, \dots$ , and if  $T = \{t_j\}$ , each  $B_k$  may be written as

$$B_k = \{x(\cdot); (x(t_1), \dots) \in D_k\}, \quad D_k \in \mathcal{B}_\infty.$$

Now it is pretty clear that any countable set combinations of the  $B_k$  produce a set with base  $T$ , hence a set in  $\mathcal{C}$ .

**Corollary 12.9.** *For  $B \in \mathcal{B}_I$ ,  $\{\omega; X(\omega) \in B\} \in \mathcal{F}$ .*

*Proof.* By the previous proposition there is a countable set  $\{t_j\}$  such that

$$B = \{x(\cdot); (x(t_1), \dots) \in D\}, \quad D \in \mathcal{B}_\infty.$$

Thus  $\{\omega; X(\omega) \in B\} = \{\omega; (X(t_1), \dots) \in D\}$ , and this is in  $\mathcal{F}$  by 2.13.

**Definition 12.10.** *The finite-dimensional distribution functions of the process are given by*

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P(X(t_1) < x_1, \dots, X(t_n) < x_n), \quad t_1 < t_2 < \dots < t_n.$$

The notation  $F_t(x)$  may also be used.

**Definition 12.11.** *The distribution of the process is the probability  $\hat{P}$  on  $\mathcal{B}$  defined by*

$$\hat{P}(B) = P(X \in B).$$

It is easy to prove that

**Proposition 12.12.** *Any two stochastic processes on  $I$  having the same finite dimensional distribution functions have the same distribution.*

*Proof.* If  $X(t), X'(t)$  are the two processes, it follows from 2.22 that  $X(t_1), X(t_2), \dots$  and  $X'(t_1), X'(t_2), \dots$  have the same distribution. Thus if  $B$  is any set with base  $\{t_1, t_2, \dots\}$ , there is a set  $D \in \mathcal{B}_\infty$  such that

$$P(X \in B) = P((X(t_1), \dots) \in D) = P'((X'(t_1), \dots) \in D) = P'(X' \in B).$$

The converse is also true, namely, that starting with a consistent set of finite dimensional distribution functions we may construct a process having those distribution functions.

**Definition 12.13.** Given a set of distribution functions

$$\{F_{t_1 \dots t_n}(x_1, \dots, x_n)\},$$

defined for all finite subsets  $\{t_1 < \dots < t_n\}$  of  $I$ . They are said to be consistent if

$$\lim_{x_k \uparrow \infty} F_{t_1 \dots t_n}(x_1, \dots, x_n) = F_{t_1 \dots t_k \dots t_n}(x_1, \dots, \hat{x}_k, \dots, x_n),$$

where the  $\hat{\phantom{x}}$  denotes missing.

**Theorem 12.14.** Given a set of consistent distribution functions as in 12.13 above, there is a stochastic process  $\{X(t)\}$ ,  $t \in I$ , such that

$$F_{t_1 \dots t_n}(x_1, \dots, x_n) = P(X(t_1) < x_1, \dots, X(t_n) < x_n).$$

*Proof.* Take  $(\Omega, \mathcal{F})$  to be  $(R^I, \mathcal{B}_I)$ . Denote by  $T, T_1, T_2$ , etc., countable subsets of  $I$ , and by  $\mathcal{B}_T$  all sets of the form

$$\{x(\cdot); (x(t_1), \dots) \in D\}, \quad D \in \mathcal{B}_\infty, \quad \{t_j\} = T.$$

By the extension theorem 2.26, there is a probability  $P_T$  on  $\mathcal{B}_T$  such that

$$P_T(x(\cdot); x(t_1) < x_1, \dots, x(t_n) < x_n) = F_{t_1 \dots t_n}(x_1, \dots, x_n).$$

Take  $B$  any set in  $\mathcal{B}_I$ , then by 12.8 there is a  $T$  such that  $B \in \mathcal{B}_T$ . We would like to define  $P$  on  $\mathcal{B}_I$  by  $P(B) = P_T(B)$ . To do this, the question is—is the definition well defined? That is, if we let  $B \in \mathcal{B}_{T_1}$ ,  $B \in \mathcal{B}_{T_2}$ , is  $P_{T_1}(B) = P_{T_2}(B)$ ? Now  $B \in \mathcal{B}_{T_1 \cup T_2}$ , hence it is sufficient to show that

$$T \subset T', B \in \mathcal{B}_T \Rightarrow P_T(B) = P_{T'}(B).$$

But  $\mathcal{B}_T \subset \mathcal{B}_{T'}$ , and  $P_T = P_{T'}$  on all rectangles with base in  $T$ ; hence  $P_{T'}$  is an extension to  $\mathcal{B}_{T'}$  of  $P_T$  on  $\mathcal{B}_T$ , so  $P_T(B) = P_{T'}(B)$ . Finally, to show  $P$  is  $\sigma$ -additive on  $\mathcal{B}_I$ , take  $\{B_n\}$  disjoint in  $\mathcal{B}_I$ ; then there is a  $T$  such that all  $B_1, B_2, \dots$  are in  $\mathcal{B}_T$ , hence so is  $\cup B_k$ . Obviously now, by the  $\sigma$ -additivity of  $P_T$ ,

$$P(\cup B_k) = P_T(\cup B_k) = \sum_k P_T(B_k) = \sum_k P(B_k).$$

The probability space is now defined. Finish by taking

$$X(t, x(\cdot)) = x(t).$$

#### 4. BEYOND THE KOLMOGOROV EXTENSION

One point of the previous paragraph was that from the definition, the most complicated sets that could be guaranteed to be in  $\mathcal{F}$  were of the form

$$\{\omega; (X(t_1), X(t_2), \dots) \in D\}, \quad D \in \mathcal{B}_\infty.$$

Furthermore, starting from the distribution functions, the extension to  $\mathcal{B}_I$  is unique, and the maximal  $\sigma$ -field to which the extension is unique is the completion,  $\bar{\mathcal{B}}_I$ , the class of all sets  $A$  such that  $A$  differs from a set in  $\mathcal{B}_I$  by a subset of a set of probability zero (see Appendix A.10).

Now consider sets of the form

$$A_1 = \{X(t, \omega) = 0 \text{ at least once for } t \in I\},$$

$$A_2 = \{|X(t)| \leq \alpha, \text{ all } t \in I\}.$$

These can be expressed as

$$A_1 = \bigcup_{t \in I} \{X(t, \omega) = 0\},$$

$$A_2 = \bigcap_{t \in I} \{|X(t, \omega)| \leq \alpha\}.$$

If each  $X(t)$  has a continuous distribution, then  $A_1$  is a noncountable union of sets of probability zero.  $A_2$  is a noncountable intersection. Neither of  $A_1$ ,  $A_2$  depends on a countable number of coordinates because

$$A_1^c = \{X(t, \omega) \neq 0 \text{ for all } t \in I\}.$$

Clearly,  $A_1^c$  does not contain any set of the form  $\{(X(t_1), \dots) \in B\}$ ,  $B \in \mathcal{B}_\infty$ . Thus  $A_1^c$  is not of the form  $\{X \in B\}$ ,  $B \in \mathcal{B}_I$ , so neither is  $A_1$ . Similarly,  $A_2$  contains no sets of the form  $\{X \in B\}$ ,  $B \in \mathcal{B}_I$ . This forces the unpleasant conclusion that if all we are given are the joint distribution functions of the process, there is no unique way of calculating the interesting and important probabilities that a process has a zero crossing during the time interval  $I$  or remains bounded below  $\alpha$  in absolute value during  $I$ . (Unless, of course, these sets accidentally fall in  $\bar{\mathcal{B}}_I$ . See Problem 3 for an important set which is not in  $\bar{\mathcal{B}}_I$ .)

But a practical approach that seems reasonable is: Let

$$A_\epsilon = \left\{ \inf_{t \in I} |X(t)| < \epsilon \right\}$$

and hope that a.s.  $A_1 = \lim_{\epsilon \downarrow 0} A_\epsilon$ . To compute  $P(A_\epsilon)$ , for  $I = [0, 1]$  say, compute

$$P_n = P\left(\inf_{k \leq n} |X(k/n)| < \epsilon\right),$$

and define  $P(A_\epsilon) = \lim_n P_n$ . Note that  $\{\inf_{k \leq n} |X(k/n)| < \epsilon\} \in \mathcal{F}$ , so its probability is well-defined and computable from the distribution functions. Finally, define  $P(A_1) = \lim_{\epsilon \downarrow 0} P(A_\epsilon)$ . This method of approximation is appealing. How to get it to make sense? We take this up in the next section.

### Problems

2. Prove for Brownian motion that the fields

$$\mathcal{F}(X(t), t \in [a, b]) \quad \text{and} \quad \mathcal{F}(X(t) - X(b), t \in [b, c]),$$

$0 \leq a \leq b < c \leq \infty$ , are independent.

3. Let

$$A = \{x(\cdot) \in R^I; x(\cdot) \text{ is } \mathcal{B}_1(I) \text{ measurable}\}.$$

Show by considering  $A$  and  $A^c$ , that  $A$  is never in  $\bar{\mathcal{B}}_I$  for any probability  $P$  on  $\mathcal{B}_I$ .

### 5. EXTENSION BY CONTINUITY

We are going to insist that all the processes we deal with in this chapter have a very weak continuity property.

**Definition 12.15.** Given a stochastic process  $\{X(t)\}$ ,  $t \in I$ , say that it is continuous in probability if for every  $t \in I$ , whenever  $t_n \rightarrow t$ , then  $X(t_n) \xrightarrow{P} X(t)$ .

When is  $X(t, \omega)$  a continuous function of  $t$ ? The difficulty here again is that the set

$$\{\omega; X(t, \omega) \text{ continuous on } I\}$$

is not necessarily in  $\mathcal{F}$ . It certainly does not depend on only a countable number of coordinates. However, one way of getting around the problem is to take  $T = \{t_j\}$  dense in  $I$ . The set

$$F = \{\omega; X(\cdot, \omega) \text{ uniformly continuous on } T\}$$

is in  $\mathcal{F}$ . To see this more clearly, for  $h > 0$  define

$$U(h, \omega) = \sup_{|t_i - t_j| \leq h} |X(t_i, \omega) - X(t_j, \omega)|, \quad t_i, t_j \in T.$$

The function  $U(h, \omega)$  is the supremum over a countable set of random variables, hence is certainly a random variable. Furthermore, it is decreasing in  $h$ . If as  $h \downarrow 0$ ,  $U(h, \omega) \xrightarrow{\text{a.s.}} 0$ , then for almost every  $\omega$ ,  $X(t, \omega)$  is a uniformly continuous function on  $T$ . Let  $C \in \mathcal{F}$  be the set on which  $U(h, \omega) \rightarrow 0$ . Assume  $P(C) = 1$ . For  $\omega \in C$ , define  $\tilde{X}(t, \omega)$  to be the unique continuous

function on  $I$  that coincides with  $X(t, \omega)$  on  $T$ . For  $t \in T$ , obviously  $\tilde{X}(t, \omega) = X(t, \omega)$ . For  $t \notin T$ ,  $\omega \in C$ ,  $\tilde{X}(t, \omega) = \lim_{t_j \rightarrow t} X(t_j, \omega)$ . Define  $\tilde{X}(t, \omega)$  to be anything continuous for  $\omega \in C^c$ , for example  $X(t, \omega) \equiv 0$ , all  $t \in I, \omega \in C^c$ . But note that  $X(t_j) \xrightarrow{\text{a.s.}} \tilde{X}(t)$  for  $t_j \in T, t_j \rightarrow t$ , which implies that  $X(t_j) \xrightarrow{P} \tilde{X}(t)$  and implies further that  $\tilde{X}(t) = X(t)$  almost surely. When  $I$  is an infinite interval, then this same construction works if for any finite interval  $J \subset I$ , there is probability one that  $X(\cdot)$  is uniformly continuous on  $T \cap J$ . Thus we have proved

**Theorem 12.16.** *If the process  $\{X(t)\}$ ,  $t \in I$ , is continuous in probability, and if there is a countable set  $T$  dense in  $I$  such that*

$$P(\omega; X(t, \omega) \text{ is uniformly continuous on } T \cap J) = 1$$

*for every finite subinterval  $J \subset I$ , then there is a process  $\tilde{X}(t, \omega)$  such that  $\tilde{X}(t, \omega)$  is a continuous function of  $t \in I$  for every fixed  $\omega$ , and for each  $t$ ,*

$$X(t, \omega) = \tilde{X}(t, \omega) \quad \text{a.s.}$$

The revised process  $\{\tilde{X}(t, \omega)\}$  and the original process  $\{X(t, \omega)\}$  have the same distribution, because for any countable  $\{t_j\}$ ,

$$P(\omega; (X(t_1), \dots) \neq (\tilde{X}(t_1), \dots)) = 0.$$

Not only have the two processes the same distribution, so that they are indistinguishable probabilistically, but the  $\{\tilde{X}(t)\}$  process is defined on the same probability space as the original process. The  $\{\tilde{X}(t)\}$  process lends itself to all the computations and devices we wanted to use before. For example: for  $I = [0, 1]$ ,

$$\tilde{A}_1 = \{\omega; \exists t \in I \text{ such that } \tilde{X}(t, \omega) = 0\},$$

$$\tilde{A}_\epsilon = \{\omega; \exists t \in I \text{ such that } |\tilde{X}(t, \omega)| < \epsilon\}.$$

It is certainly now true that

$$\tilde{A}_1 = \lim_{\epsilon \downarrow 0} \tilde{A}_\epsilon.$$

But take  $\tilde{A}_{n,\epsilon}$  to be the set  $\tilde{A}_{n,\epsilon} = \{\omega; \exists k \leq 2^n, \text{ such that } |\tilde{X}(k/2^n)| < \epsilon\}$ . Then

$$\tilde{A}_{n,\epsilon} \in \mathcal{F}, \tilde{A}_{n,\epsilon} \uparrow \tilde{A}_\epsilon,$$

which implies  $\tilde{A}_\epsilon \in \mathcal{F}$ , so, in turn,  $\tilde{A}_1 \in \mathcal{F}$ . Furthermore,

$$P(\tilde{A}_\epsilon) = \lim_n P(\tilde{A}_{n,\epsilon}).$$

Therefore, by slightly revising the original process, we arrive at a process having the same distribution, for which the reasonable approximation procedures we wish to use are valid and the various interesting sets are measurable.

Obviously not all interesting stochastic processes can be altered slightly so as to have all sample functions continuous. But the basic idea always is to pick and work with the smoothest possible version of the process.

**Definition 12.17.** Given two processes  $\{X(t)\}$  and  $\{\tilde{X}(t)\}$ ,  $t \in I$ , on the same probability space  $(\Omega, \mathcal{F}, P)$ . They will be called versions of each other if

$$P(X(t) \neq \tilde{X}(t)) = 0, \quad \text{all } t \in I.$$

### Problems

4. Show that if a process  $\{X(t)\}$ ,  $t \in I$ , is continuous in probability, then for any set  $\{t_n\}$  dense in  $I$ , each  $X(t)$  is measurable with respect to the completion of  $\mathcal{F}(X(t_1), X(t_2), \dots)$ , or that each set of  $\mathcal{F}(X(t), t \in I)$  differs from a set of  $\mathcal{F}(X(t_1), \dots)$  by a set of probability zero.

5. Conclude from the above that if  $T_n \subset T_{n+1}$ ,  $T_n \uparrow T$ ,  $T_n$  finite subsets of  $I$ ,  $T$  dense in  $I$ , then for  $J \subset I$ ,  $A \in \mathcal{F}$ ,

$$P(A | X(t), t \in J) = \lim_{n \rightarrow \infty} P(A | X(t), t \in J \cap T_n), \quad \text{a.s.}$$

6. If  $X(t)$ ,  $\tilde{X}(t)$  are versions of each other for  $t \in I$ , and if both processes have all sample paths continuous on  $I$ , show that

$$P(X(t) = \tilde{X}(t) \text{ for all } t \in I) = 1.$$

7. If  $\{X(t)\}$ ,  $t \in I$ , is a process all of whose paths are continuous on  $I$ , then show that the function  $X(t, \omega)$  defined on  $I \times \Omega$  is measurable with respect to the product  $\sigma$ -field  $\mathcal{B}_1(I) \times \mathcal{F}$ . [For  $I$  finite, let  $I_1, \dots, I_n$  be any partition of  $I$ ,  $t_k \in I_k$ , and consider approximating  $X(t)$  by the functions  $\sum_k X(t_k) \chi_{I_k}(t)$ .]

### 6. CONTINUITY OF BROWNIAN MOTION

It is easy to check that the finite-dimensional distributions given by 12.5 are consistent. Hence there is a process  $\{X(t)\}$ ,  $t \geq 0$  fitting them.

**Definition 12.18.** Let  $X(t)$  be a Brownian motion. If  $\mu \neq 0$  it is said to be a Brownian motion with drift  $\mu$ . If  $\mu = 0$ ,  $\sigma^2 = 1$ , it is called normalized Brownian motion, or simply Brownian motion.

Note that  $(X(t) - \mu t)/\sigma$  is normalized. The most important single sample path property is contained in

**Theorem 12.19.** For any Brownian motion  $X(t)$  there is a dense set  $T$  in  $[0, \infty)$  such that  $X(t)$  is uniformly continuous on  $T \cap [0, a]$ ,  $a < \infty$ , for almost every  $\omega$ .

In preparation for the proof of this, we need

**Proposition 12.20.** *Let  $T_0$  be any finite collection of points,  $0 = t_0 < t_1 < \dots < t_n = \tau$ ; then for  $X(t)$  normalized Brownian motion*

$$P\left(\max_{t \in T_0} X(t) > x\right) \leq 2P(X(\tau) > x),$$

$$P\left(\max_{t \in T_0} |X(t)| > x\right) \leq 2P(|X(\tau)| > x).$$

*Proof.* Denote  $Y_k = X(t_k) - X(t_{k-1})$ , and  $j^* = \{\text{first } j \text{ such that } X(t_j) > x\}$ . Then because  $X(t_n) - X(t_j)$  has a distribution symmetric about the origin,

$$\frac{1}{2}P\left(\max_{t \in T_0} X(t) > x\right) = \sum_{j=1}^n P(j^* = j)P(X(t_n) - X(t_j) > 0).$$

Since  $\{j^* = j\} \in \mathcal{F}(X(t_0), \dots, X(t_j))$ , then

$$\begin{aligned} \frac{1}{2}P\left(\max_{t \in T_0} X(t) > x\right) &= \sum_{j=1}^n P(j^* = j, X(t_n) - X(t_j) > 0) \\ &\leq \sum_{j=1}^n P(j^* = j, X(t_n) > x) \leq P(X(\tau) > x). \end{aligned}$$

For the second inequality, use

$$\left\{\max_{t \in T_0} |X(t)| > x\right\} = \left\{\max_{t \in T_0} X(t) > x\right\} \cup \left\{\min_{t \in T_0} X(t) < -x\right\}$$

and the fact that  $-X(t)$  is normalized Brownian motion.

*Proof of 12.19.* We show this for  $a = 1$ . Take  $T_n = \{k/2^n; k = 0, \dots, 2^n\}$ , and  $T = \bigcup_1^\infty T_n$ . Define

$$U_n = \sup_{\substack{t_j, t_k \in T \\ |t_j - t_k| \leq 1/2^n}} |X(t_j) - X(t_k)|.$$

To show that  $U_n \rightarrow 0$  a.s., since  $U_n \in \downarrow$  it is sufficient to show that

$$\lim_n P(U_n > \delta) = 0.$$

Let  $I_k = [k/2^n, (k+1)/2^n]$ , and

$$Y_k = \sup_{t \in I_k \cap T} |X(t) - X(k/2^n)|, \quad k = 0, 1, \dots, 2^n - 1.$$

By the triangle inequality,

$$U_n \leq 3 \max_k Y_k.$$

We show that  $P\left(\max_k Y_k > \delta\right) \rightarrow 0$ . Use

$$P\left(\max_k Y_k > \delta\right) = P\left(\bigcup_k \{Y_k > \delta\}\right) \leq \sum_0^{2^n-1} P(Y_k > \delta).$$

The  $Y_1, \dots$  are identically distributed, so

$$P\left(\max_k Y_k > \delta\right) \leq 2^n P(Y_0 > \delta).$$

Note that

$$Y_0 = \lim_N \max_{t \in I_0 \cap T_N} |X(t)|.$$

By 12.20, for  $N \geq n$ ,

$$P\left(\max_{t \in I_0 \cap T_N} |X(t)| > \delta\right) \leq 2P(|X(2^{-n})| > \delta),$$

hence

$$P(Y_0 > \delta) \leq 2P(|X(2^{-n})| > \delta).$$

Since Brownian motion satisfies  $P(|X(\Delta t)| > \delta)/\Delta t \rightarrow 0$  as  $\Delta t \rightarrow 0$ , then

$$2^n P(|X(2^{-n})| > \delta) \rightarrow 0,$$

which proves the theorem.

**Corollary 12.21.** *There is a version of Brownian motion on  $[0, \infty)$  such that all sample paths are continuous.*

Henceforth, we assume that the Brownian motion we deal with has all sample paths continuous.

## Problems

8. Prove that for  $\{X(t)\}$  normalized Brownian motion on  $[0, \infty)$ ,  $P(X(n\delta) \in J \text{ i.o.}) = 1$  for all intervals  $J$  such that  $\|J\| > 0$ , and fixed  $\delta > 0$ .
9. Define  $\mathcal{G} = \bigcap_{t \geq 0} \mathcal{F}(X(\tau), \tau \geq t)$ , or  $A \in \mathcal{G}$  if  $A \in \mathcal{F}(X(\tau), \tau \geq t)$  for all  $t \geq 0$ . Prove that  $A \in \mathcal{G}$  implies  $P(A) = 0$  or 1. [Apply a generalized version of 3.50.]

## 7. AN ALTERNATIVE DEFINITION

Normalized Brownian motion is completely specified by stating that it is a Gaussian process, i.e., all finite subsets  $\{X(t_1), \dots, X(t_n)\}$  have a joint normal distribution,  $EX(t) = 0$  and covariance  $\Gamma(s, t) = \min(s, t)$ . Since all sample functions are continuous, to specify Brownian motion it would only be necessary to work with a countable subset of random variables



$\{X(t)\}$ ,  $t \in T$ , and get the others as limits. This leads to the idea that with a proper choice of coordinates,  $X(t)$  could be expanded in a countable coordinate system. Let  $Y_1, Y_2, \dots$  be independent and  $N(0, 1)$ . Let  $\varphi_k(t)$  be defined on a closed interval  $I$  such that  $\sum_1^\infty |\varphi_k(t)|^2 < \infty$ , all  $t \in I$ . Consider

$$Z(t) = \sum_1^\infty \varphi_k(t) Y_k.$$

Since

$$\sum_1^\infty E |\varphi_k(t) Y_k|^2 = \sum_1^\infty |\varphi_k(t)|^2 < \infty,$$

the sums converge a.s. for every  $t \in I$ , hence  $Z(t)$  is well-defined for each  $t$  except on a set of probability zero. Furthermore,  $Z(t_1), \dots, Z(t_n)$  is the limit in distribution of joint normal variables, so that  $\{Z(t)\}$  is a Gaussian process.

Note  $EZ(t) = 0$ , and for the  $Z(t)$  covariance

$$(12.22) \quad \Gamma(s, t) = \sum_1^\infty \varphi_k(s) \varphi_k(t), \quad (s, t) \in I \times I.$$

Hence if the  $\varphi_k(t)$  satisfy

$$\min(s, t) = \sum_1^\infty \varphi_k(s) \varphi_k(t), \quad (s, t) \in I \times I,$$

then  $Z(t)$  is normalized Brownian motion on  $I$ . I assert that on  $I = [0, \pi]$ ,

$$(12.23) \quad \min(s, t) = \frac{ts}{\pi} + \frac{2}{\pi} \sum_{k \geq 1} \frac{\sin kt \sin ks}{k^2}.$$

One way to verify this is to define a function of  $t$  on  $[-\pi, \pi]$  for any  $s \geq 0$ , by

$$h_s(t) = \begin{cases} t, & |t| \leq s; \\ -s, & t \leq -s; \\ s, & t \geq s. \end{cases}$$

Denote the right-hand side of (12.23) by  $g_s(t)$ . The sum converges uniformly, hence  $g_s(t)$  is continuous for all  $s, t$ . Simply check that for all integers  $k$ ,

$$\int_{[-\pi, +\pi]} e^{ikt} h_s(t) dt = \int_{[-\pi, +\pi]} e^{ikt} g_s(t) dt,$$

and use the well-known fact that two continuous functions with the same Fourier coefficients are equal on  $[-\pi, +\pi]$ . Since  $h_s(t) = \min(s, t)$ , for  $t \geq 0$ , (12.23) results.

**Proposition 12.24.** Let  $Y_0, Y_1, \dots$  be independent  $N(0, 1)$ , then

$$X(t) = \frac{t}{\sqrt{\pi}} Y_0 + \sqrt{\frac{2}{\pi}} \sum_{m=1}^{\infty} \frac{\sin mt}{m} Y_m$$

is normalized Brownian motion on  $[0, \pi]$ .

One way to prove the continuity of sample paths would be to define  $X^{(n)}(t)$  as the  $n$ th partial sum in 12.24, and to show that for almost all  $\omega$ , the functions  $x_n(t) = X^{(n)}(t, \omega)$  converged uniformly on  $[0, \pi]$ . This can be shown true, at least for a subsequence  $X^{(n')}(t)$ . See Ito and McKean [76, p. 22], for a proof along these lines.

## 8. VARIATION AND DIFFERENTIABILITY

The Brownian motion paths are extremely badly behaved for continuous functions. Their more obvious indices of bad behavior are given in this section: they are nowhere differentiable, and consequently of unbounded variation in every interval.

**Theorem 12.25.** Almost every Brownian path is nowhere differentiable.

*Proof.* We follow Dvoretzki, Erdős, and Kakutani [42]. Fix  $\beta > 0$ , suppose that a function  $x(t)$  has derivative  $x'(s)$ ,  $|x'(s)| < \beta$ , at some point  $s \in [0, 1]$ ; then there is an  $n_0$  such that for  $n > n_0$

$$(12.26) \quad |x(t) - x(s)| \leq 2\beta |t - s|, \quad \text{if } |t - s| \leq 2/n.$$

Let  $x(\cdot)$  denote functions on  $[0, 1]$ .

$$A_n = \{x(\cdot); \exists s \text{ such that } |x(t) - x(s)| \leq 2\beta |t - s|, \text{ if } |t - s| \leq 2/n\}.$$

The  $A_n$  increase with  $n$ , and the limit set  $A$  includes the set of all sample paths on  $[0, 1]$  having a derivative at any point which is less than  $\beta$  in absolute value. If (12.26) holds, then, and we let  $k$  be the largest integer such that  $k/n \leq s$ , the following is implied:

$$\begin{aligned} y_k &= \\ &= \max \left( \left| x\left(\frac{k+2}{n}\right) - x\left(\frac{k+1}{n}\right) \right|, \left| x\left(\frac{k+1}{n}\right) - x\left(\frac{k}{n}\right) \right|, \left| x\left(\frac{k}{n}\right) - x\left(\frac{k-1}{n}\right) \right| \right) \\ &\leq \frac{6\beta}{n}. \end{aligned}$$

Therefore, if

$$B_n = \left\{ x(\cdot); \text{ at least one } y_k \leq \frac{6\beta}{n} \right\},$$

then  $A_n \subset B_n$ . Thus to show  $P(A) = 0$ , which implies the theorem, it is sufficient to get  $\lim_n P(B_n) = 0$ . But

$$\begin{aligned}
 B_n &= \bigcup_{k=1}^{n-2} \left\{ x(\cdot); \quad y_k \leq \frac{6\beta}{n} \right\}, \\
 P(B_n) &\leq \sum_1^{n-2} P \left( \max \left( \left| x\left(\frac{k+2}{n}\right) - x\left(\frac{k+1}{n}\right) \right|, \left| x\left(\frac{k+1}{n}\right) - x\left(\frac{k}{n}\right) \right|, \right. \right. \\
 &\quad \left. \left. \left| x\left(\frac{k}{n}\right) - x\left(\frac{k-1}{n}\right) \right| \right) \leq \frac{6\beta}{n} \right) \\
 &\leq nP \left( \max \left( \left| x\left(\frac{3}{n}\right) - x\left(\frac{2}{n}\right) \right|, \left| x\left(\frac{2}{n}\right) - x\left(\frac{1}{n}\right) \right|, \right. \right. \\
 &\quad \left. \left. \left| x\left(\frac{1}{n}\right) \right| \right) \leq \frac{6\beta}{n} \right) \\
 &= nP^3 \left( \left| x\left(\frac{1}{n}\right) \right| \leq \frac{6\beta}{n} \right) = n \left[ \sqrt{\frac{n}{2\pi}} \int_{-6\beta/n}^{+6\beta/n} e^{-nx^2/2} dx \right]^3.
 \end{aligned}$$

Substitute  $nx = y$ . Then

$$P(B_n) \leq n \left[ \frac{1}{\sqrt{2\pi n}} \int_{-6\beta}^{+6\beta} e^{-x^2/2n} dx \right]^3 \rightarrow 0.$$

**Corollary 12.27.** *Almost every sample path of  $X(t)$  has infinite variation on every finite interval.*

*Proof.* If a sample function  $X(t, \omega)$  has bounded variation on  $I$ , then it has a derivative existing almost everywhere on  $I$ .

A further result gives more information on the size of the oscillations of  $X(t)$ . Since  $E |X(t + \Delta t) - X(t)|^2 = \Delta t$ , as a rough estimate we would guess that  $|X(t + \Delta t) - X(t)| \simeq \sqrt{\Delta t}$ . Then for any fine partition  $t_0, \dots, t_n$  of the interval  $[t, t + \tau]$ ,

$$\sum_1^n |X(t_k) - X(t_{k-1})|^2 \simeq \tau.$$

The result of the following theorem not only verifies this, but makes it surprisingly precise.

**Theorem 12.28.** *Let the partitions  $\mathcal{T}_n$  of  $[t, t + \tau]$ ,  $\mathcal{T}_n = (t_0^{(n)}, \dots, t_m^{(n)})$ ,  $\|\mathcal{T}_n\| = \sup_k |t_{k+1}^{(n)} - t_k^{(n)}|$  satisfy  $\|\mathcal{T}_n\| \rightarrow 0$ . Then*

$$S_n = \sum_k |X(t_k^{(n)}) - X(t_{k-1}^{(n)})|^2 \xrightarrow{2} \tau.$$

If  $\sum_1^\infty \|\mathcal{T}_n\| < \infty$ ,  $S_n \xrightarrow{\text{a.s.}} \tau$ .

*Proof.* Assume  $t = t_0^{(n)}$ ,  $t + \tau = t_m^{(n)}$ , otherwise do some slight modification. Then  $\tau = \sum_1^m (t_k - t_{k-1})$  (dropping the superscript), and

$$\begin{aligned} S_n - \tau &= \sum_1^m [(X(t_k) - X(t_{k-1}))^2 - (t_k - t_{k-1})] \\ &= \sum_1^m [(X(t_k) - X(t_{k-1}))^2 - E(X(t_k) - X(t_{k-1}))^2]. \end{aligned}$$

The summands are independent, with zero means. Hence

$$E(S_n - \tau)^2 = \sum_1^m E[(X(t_k) - X(t_{k-1}))^2 - (t_k - t_{k-1})]^2.$$

$(X(t_k) - X(t_{k-1}))^2 / (t_k - t_{k-1})$  has the distribution of  $X^2$ , where  $X$  is  $N(0, 1)$ . So

$$\begin{aligned} E(S_n - \tau)^2 &= E(X^2 - 1)^2 \cdot \sum_1^m (t_k - t_{k-1})^2 \\ &\leq E(X^2 - 1)^2 \cdot \|\mathcal{F}_n\| \tau, \end{aligned}$$

proving convergence in mean square. If  $\sum \|\mathcal{F}_n\| < \infty$ , then use the Borel-Cantelli lemma plus the Chebyshev inequality.

Theorem 12.28 holds more generally, with  $S_n \xrightarrow{\text{a.s.}} \tau$  for any sequence  $\mathcal{F}_n$  of partitions such that  $\|\mathcal{F}_n\| \rightarrow 0$  and the  $\mathcal{F}_n$  are successive refinements. (See Doob [39, pp. 395 ff.] )

## 9. LAW OF THE ITERATED LOGARITHM

Now for one of the most precise and well-known theorems regarding oscillations of a Brownian motion. It has gone through many refinements and generalizations since its proof by Khintchine in 1924.

**Theorem 12.29.** *For normalized Brownian motion*

$$\lim_{t \downarrow 0} \frac{X(t)}{\sqrt{2t \log(\log 1/t)}} = 1 \quad \text{a.s.}$$

*Proof.* I follow essentially Lévy's proof. Let  $\varphi(t) = \sqrt{2t \log(\log 1/t)}$ .

1) For any  $\delta > 0$ ,

$$P(\lim_{t \downarrow 0} (X(t)/\varphi(t)) > 1 + \delta) = 0.$$

*Proof of (1).* Take  $q$  any number in  $(0, 1)$ , put  $t_n = q^n$ .

The plan is to show that if  $C_n$  is the event

$$\{X(t) > (1 + \delta)\varphi(t) \text{ for at least one } t \in [t_{n+1}, t_n]\},$$

then  $P(C_n \text{ i.o.}) = 0$ . Define  $M(\tau) = \sup_{t \leq \tau} X(t)$  and use

$$C_n \subset \{M(t_n) > (1 + \delta)\varphi(t_{n+1})\},$$

valid since  $\varphi(t)$  is increasing in  $t$ . Use the estimates gotten from taking limits in 12.20.

$$\begin{aligned} P(M(t_n) > x\sqrt{t_n}) &\leq 2P(X(t_n) > x\sqrt{t_n}) = \sqrt{\frac{2}{\pi}} \int_x^\infty e^{-z^2/2} dz \\ &\leq \sqrt{\frac{2}{\pi}} \frac{1}{x} \int_x^\infty z e^{-z^2/2} dz = \sqrt{\frac{2}{\pi}} \frac{e^{-x^2/2}}{x}. \end{aligned}$$

Hence, letting  $x_n = (1 + \delta)\varphi(t_{n+1})/\sqrt{t_n}$

$$P(M(t_n) > x_n\sqrt{t_n}) \leq \sqrt{\frac{2}{\pi}} \frac{\exp[-x_n^2/2]}{x_n}.$$

and since

$$\begin{aligned} x_n &= (1 + \delta)\sqrt{2q \log[(n+1) \log 1/q]} \\ &= \sqrt{2 \log c(n+1)^\lambda}, \end{aligned}$$

where  $\lambda = q(1 + \delta)^2$ ,

$$\begin{aligned} (12.30) \quad P(M(t_n) > (1 + \delta)\varphi(t_{n+1})) \\ \leq c_1 \cdot \frac{1}{(n+1)^\lambda \sqrt{\log(n+1)}}. \end{aligned}$$

For any  $\delta$ , select  $q$  so that  $q(1 + \delta)^2 > 1$ . Then the right-hand side of (12.30) is a term of a convergent sum and the first assertion is proved.

2) For any  $\delta > 0$ ,

$$P(\lim_{t \downarrow 0} (X(t)/\varphi(t)) > 1 - \delta) = 1.$$

*Proof of (2).* Take  $q$  again in  $(0, 1)$ ,  $t_n = q^n$ , let  $Z_n = X(t_n) - X(t_{n+1})$ . The  $Z_n$  are independent. Suppose we could show that for  $\epsilon > 0$ ,

$$P(Z_n > (1 - \epsilon)\varphi(t_n) \text{ i.o.}) = 1.$$

This would be easy in principle, because the independence of the  $Z_n$  allows the converse half of the Borel-Cantelli lemma to be used. On the other hand, from part one of this proof, because the processes  $\{X(t)\}$  and  $\{-X(t)\}$  have the same distribution,

$$P(X(t_{n+1}) < -(1 + \epsilon)\varphi(t_{n+1}) \text{ i.o.}) = 0$$

or

$$X(t_{n+1}) \geq -(1 + \epsilon)\varphi(t_{n+1})$$

holds for all  $n$  sufficiently large. From  $X(t_n) = Z_n + X(t_{n+1})$  it follows that infinitely often

$$\begin{aligned} X(t_n) &\geq (1 - \epsilon)\varphi(t_n) - (1 + \epsilon)\varphi(t_{n+1}) \\ &= \varphi(t_n) \left[ 1 - \epsilon - (1 + \epsilon) \frac{\varphi(t_{n+1})}{\varphi(t_n)} \right], \quad \text{a.s.} \end{aligned}$$

Note that  $\varphi(t_{n+1})/\varphi(t_n) \rightarrow \sqrt{q}$ . Therefore, if we take  $\epsilon, q$  so small that

$$1 - \epsilon - 2(1 + \epsilon)\sqrt{q} \geq 1 - \delta,$$

the second part would be established. So now, we start estimating:

$$P(Z_n > x\sqrt{(t_n - t_{n+1})}) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-z^2/2} dz \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$$

as  $x \rightarrow \infty$ . Let

$$\begin{aligned} x_n &= (1 - \epsilon) \frac{\varphi(t_n)}{\sqrt{t_n - t_{n+1}}} = \frac{(1 - \epsilon)}{\sqrt{1 - q}} \sqrt{2 \log(n \log 1/q)} \\ &= \sqrt{\alpha \log cn}, \quad \alpha = 2(1 - \epsilon)^2/(1 - q). \end{aligned}$$

Then

$$P(Z_n > (1 - \epsilon)\varphi(t_n)) \sim \frac{c}{\sqrt{2\pi} n^{\alpha/2} \sqrt{\log n}}.$$

By taking  $q$  even smaller, if necessary, we can get  $\alpha < 2$ . The right-hand side is then a term of a divergent series and the proof is complete. Q.E.D.

## 10. BEHAVIOR AT $t = \infty$

Let  $Y_k = X(k) - X(k - 1)$ . The  $Y_k$  are independent  $\mathcal{N}(0, 1)$  variables,  $X(n) = Y_1 + \cdots + Y_n$  is the sum of independent, identically distributed random variables. Thus  $X(t)$  for  $t$  large has the magnitude properties of successive sums  $S_n$ . In particular,

### Proposition 12.31

$$\frac{X(t)}{t} \xrightarrow{\text{a.s.}} 0 \quad \text{as } t \rightarrow \infty.$$

*Proof.* Since  $EY_k = 0$ , we can use the strong law of large numbers to get  $X(n)/n \xrightarrow{\text{a.s.}} 0$ . Let

$$Z_k = \max_{0 \leq t \leq 1} |X(k + t) - X(k)|.$$

For  $t \in [k, k + 1]$ ,

$$\left| \frac{X(t)}{t} - \frac{X(k)}{k} \right| \leq \frac{1}{k(k + 1)} |X(k)| + \frac{1}{k + 1} Z_k.$$

The first term  $\xrightarrow{\text{a.s.}} 0$ .  $Z_k$  has the same distribution as  $\max_{0 \leq t \leq 1} |X(t)|$ . By 12.20,  $EZ_k < \infty$ . Now use Problem 10, Chapter 3, to conclude that  $Z_k/k \xrightarrow{\text{a.s.}} 0$ .

This is the straightforward approach. There is another way which is surprising, because it essentially reduces behavior for  $t \rightarrow \infty$  to behavior for  $t \downarrow 0$ . Define

$$\begin{aligned} X^{(1)}(t) &= tX(1/t), \quad t > 0, \\ &\equiv 0, \quad t = 0. \end{aligned}$$

**Proposition 12.32.**  $X^{(1)}(t)$  is normalized Brownian motion on  $[0, \infty)$ .

*Proof.* Certainly  $X^{(1)}(t)$  is Gaussian with zero mean. Also,

$$EX^{(1)}(t)X^{(1)}(s) = ts \min\left(\frac{1}{t}, \frac{1}{s}\right) = \min(s, t).$$

Now to prove 12.31 another way. The statement  $X(t)/t \rightarrow 0$  as  $t \rightarrow \infty$  translates into  $tX(1/t) \rightarrow 0$  a.s. as  $t \rightarrow 0$ . So 12.31 is equivalent to proving that  $X^{(1)}(t) \rightarrow 0$  a.s. as  $t \rightarrow 0$ . If  $X^{(1)}(t)$  is a version of Brownian motion with all paths continuous on  $[0, \infty)$ , then trivially,  $X^{(1)}(t) \rightarrow 0$  a.s. at the origin. However, the continuity of  $X(t)$  on  $[0, \infty)$  gives us only that the paths of  $X^{(1)}(t)$  are continuous on  $(0, \infty)$ . Take a version  $\tilde{X}^{(1)}(t)$  of  $X^{(1)}(t)$  such that all paths of  $\tilde{X}^{(1)}(t)$  are continuous on  $[0, \infty)$ . By Problem 5, almost all paths of  $\tilde{X}^{(1)}(t)$  and  $X^{(1)}(t)$  coincide on  $(0, \infty)$ . Since  $\tilde{X}^{(1)}(t) \rightarrow 0$  as  $t \rightarrow 0$ , this is sufficient.

By using this inversion on the law of the iterated logarithm we get

**Corollary 12.33**

$$\overline{\lim}_{t \rightarrow \infty} \frac{X(t)}{\sqrt{2t \log(\log t)}} = 1.$$

Since  $-X(t)$  is also Brownian motion,

$$\lim_{t \rightarrow \infty} \frac{X(t)}{\sqrt{2t \log(\log t)}} = -1.$$

Therefore,

$$\overline{\lim}_{t \rightarrow \infty} \frac{|X(t)|}{\sqrt{2t \log(\log t)}} = 1.$$

The similar versions of 12.29 hold as  $t \rightarrow 0$ ; for instance,

$$(12.34) \quad \lim_{t \rightarrow 0} \frac{X(t)}{\sqrt{2t \log |\log t|}} = -1.$$

11. THE ZEROS OF  $X(t)$ 

Look at the set  $T(\omega)$  of zeros of  $X(t, \omega)$  in the interval  $[0, 1]$ . For any continuous function, the zero set is closed. By (12.29) and (12.34),  $T(\omega)$  is an infinite set a.s. Furthermore the Lebesgue measure of  $T(\omega)$  is a.s. zero, because  $l(T(\omega)) = l\{t; X(t) = 0\} = \int_0^1 \chi_{\{0\}}(X(t)) dt$ , so

$$El(T(\omega)) = E \int_0^1 \chi_{\{0\}}(X(t)) dt = \int_0^1 P(X(t) = 0) dt = 0,$$

where the interchange of  $E$  and  $\int_0^1 dt$  is justified by the joint measurability of  $X(t, \omega)$ , hence of  $\chi_{\{0\}}(X(t, \omega))$ . (See Problem 7.)

**Theorem 12.35.** *For almost all  $\omega$ ,  $T(\omega)$  is a closed, perfect set of Lebesgue measure zero (therefore, noncountable).*

*Proof.* The remaining part is to prove that  $T(\omega)$  has no isolated points. The idea here is that every time  $X(t)$  hits zero, it is like starting all over again and the law of the iterated logarithm guarantees a clustering of zeros starting from that point. For almost all paths, the point  $t = 0$  is a limit point of zeros of  $X(t)$  from the right. For any point  $a \geq 0$ , let  $\tau^*$  be the position of the first zero of  $X(t)$  following  $t = a$ , that is,

$$\tau^* = \inf \{t; X(t) = 0, t \geq a\}.$$

Look at the process  $X^{(1)}(t) = X(t + \tau^*) - X(\tau^*)$ . This is just looking at the Brownian process as though it started afresh at the time  $\tau^*$ . Heuristically, what happens up to time  $\tau^*$  depends only on the process up to that time; starting over again at  $\tau^*$  should give a process that looks exactly like Brownian motion. If this argument can be made rigorous, then the set  $C_a$  of sample paths such that  $\tau^*$  is a limit point of zeros from the right has probability one. The intersection of  $C_a$  over all rational  $a \geq 0$  has probability one, also. Therefore almost every sample path has the property that the first zero following any rational is a limit point of zeros from the right. This precludes the existence of any isolated zero. Therefore, the theorem is proved except for the assertion,

(12.36)  $X(t + \tau^*) - X(\tau^*)$  is a Brownian motion.

The truth of (12.36) and its generalization are established in the next section. Suppose that it holds for more general random starting times. Then we could use this to prove

**Corollary 12.37.** *For any value  $a$ , the set  $T^{(a)} = \{t; X(t) = a, 0 \leq t \leq 1\}$  is, for almost all  $\omega$ , either empty or a perfect closed set of Lebesgue measure zero.*



*Proof.* Let  $t^*$  be the first  $t$  such that  $X(t) = a$ , that is,  $t^* = \inf \{t; X(t) = a\}$ . If  $t^* > 1$ , then  $T^{(a)}$  is empty. The set  $\{t^* = 1\} \subset \{X(1) = a\}$  has probability zero. If  $t^* < 1$ , consider the process  $X(t + t^*) - X(t^*)$  as starting out at the random time  $t^*$ . If this is Brownian motion, the zero set is perfect. But the zero set for this process is

$$\begin{aligned} T &= \{t; X(t + t^*) - X(t^*) = 0, \quad 0 \leq t \leq 1\} \\ &= \{t; X(t + t^*) = a, \quad 0 \leq t \leq 1\}. \end{aligned}$$

Hence  $T^{(a)} = T \cap [0, 1]$  is perfect a.s.

## 12. THE STRONG MARKOV PROPERTY

The last item needed to complete and round out this study of sample path properties is a formulation and proof of the statement: At a certain time  $t^*$ , where  $t^*$  depends only on the Brownian path up to time  $t^*$ , consider the motion as starting at  $t^*$ ; that is, look at  $X^{(1)}(t) = X(t + t^*) - X(t^*)$ . Then  $X^{(1)}(t)$  is Brownian motion and is independent of the path of the particle up to time  $t^*$ . Start with the observation that for  $\tau \geq 0$ , fixed,

$$(12.38) \quad X^{(1)}(t) = X(t + \tau) - X(\tau), \quad t \geq 0$$

is Brownian motion and is independent of  $\mathcal{F}(X(s), s \leq \tau)$ . Now, to have any of this make sense, we need:

**Proposition 12.39.** *If  $t^* \geq 0$  is a random variable so is  $X(t^*)$ .*

*Proof.* For any  $n > 0$ , let

$$\begin{aligned} A_k^{(n)} &= \left\{ \frac{k-1}{n} \leq t^* < \frac{k}{n} \right\}, \\ X^{(n)}(t^*) &= \sum_{k=1}^{\infty} X\left(\frac{k}{n}\right) \chi_{A_k^{(n)}}. \end{aligned}$$

$X^{(n)}(t^*)$  is a random variable. On the set  $\{t^* \leq N\}$ ,

$$|X^{(n)}(t^*) - X(t^*)| \leq \sup_{\substack{0 \leq t \leq N \\ 0 \leq h \leq 1/n}} |X(t+h) - X(t)|.$$

The right-hand side  $\rightarrow 0$ , so  $X^{(n)}(t^*) \rightarrow X(t^*)$  everywhere.

Next, it is necessary to formulate the statement that the value of  $t^*$  depends only on the past of the process up to time  $t^*$ .

**Definition 12.40.** *For any process  $\{X(t)\}$  a random variable  $t^* \geq 0$  will be called a stopping time if for every  $t \geq 0$ ,*

$$\{t^* \leq t\} \in \mathcal{F}(X(\tau), \tau \leq t).$$

The last step is to give meaning to “the part of the process up to time  $t^*$ .”

Look at an example:

$$\tau^* = \inf_t \{t; X(t) \geq 1\},$$

so  $\tau^*$  is the first time that  $X(t)$  hits the point  $x = 1$ . It can be shown quickly that  $\tau^*$  is a stopping time. Look at sets depending on  $X(t)$ ,  $0 \leq t \leq \tau^*$ ; for example,

$$B = \left\{ \omega; \inf_{t \leq \tau^*} X(t) \leq -1 \right\}.$$

Note that for any  $t \geq 0$ ,  $B \cap \{\tau^* \leq t\}$  depends only on the behavior of the sample path on  $[0, t]$ , that is,

$$B \cap \{\tau^* \leq t\} \in \mathcal{F}(X(\tau), \tau \leq t).$$

Generalizing from this we get

**Definition 12.41.** The  $\sigma$ -field of events  $B \in \mathcal{F}$  such that for every  $t \geq 0$ ,  $B \cap \{\tau^* \leq t\} \in \mathcal{F}(X(\tau), \tau \leq t)$  is called the  $\sigma$ -field generated by the process up to time  $\tau^*$  and denoted by  $\mathcal{F}(X(t), t \leq \tau^*)$ .

This is all we need:

**Theorem 12.42.** Let  $\tau^*$  be a stopping time, then

$$X^{(1)}(t) = X(t + \tau^*) - X(\tau^*), \quad t \geq 0$$

is normalized Brownian motion and

$$\mathcal{F}(X^{(1)}(t), t \geq 0) \quad \text{is independent of} \quad \mathcal{F}(X(t), t \leq \tau^*).$$

*Proof.* If  $\tau^*$  takes on only a countable number of values  $\{\tau_k\}$ , then 12.42 is quick. For example, if

$$A_1, \dots, A_j \in \mathcal{B}_1, \quad t_1, \dots, t_j \geq 0, \quad \text{and} \quad B \in \mathcal{F}(X(t), t \leq \tau^*),$$

then

$$\begin{aligned} P(X^{(1)}(t_1) \in A_1, \dots, X^{(1)}(t_j) \in A_j, B) \\ = \sum_k P(X^{(1)}(t_1) \in A_1, \dots, X^{(1)}(t_j) \in A_j, \tau^* = \tau_k, B). \end{aligned}$$

Now,

$$\{X^{(1)}(t_1) \in A_1, \dots, \tau^* = \tau_k, B\} = \{X(t_1 + \tau_k) - X(\tau_k) \in A_1, \dots, \tau^* = \tau_k, B\}.$$

Furthermore,

$$\{\tau^* = \tau_k\} \cap B = \{\tau^* = \tau_k\} \cap \{\tau^* \leq \tau_k\} \cap B,$$

so is in  $\mathcal{F}(X(t), t \leq \tau_k)$ . By (12.38), then,

$$\begin{aligned} P(X^{(1)}(t_1) \in A_1, \dots, \mathbf{t}^* = \tau_k, B) \\ = P(X(t_1 + \tau_k) - X(\tau_k) \in A_1, \dots, X(t_j + \tau_k) - X(\tau_k) \in A_j) \\ \cdot P(\mathbf{t}^* = \tau_k, B) \\ = P(X(t_1) \in A_1, \dots, X(t_j) \in A_j) P(\mathbf{t}^* = \tau_k, B). \end{aligned}$$

Summing over  $k$ , we find that

$$\begin{aligned} (12.43) \quad P(X^{(1)}(t_1) \in A_1, \dots, X^{(1)}(t_j) \in A_j, B) \\ = P(X(t_1) \in A_1, \dots, X(t_j) \in A_j) P(B). \end{aligned}$$

This extends immediately to the statement of the theorem. In the general case, we approximate  $\mathbf{t}^*$  by a discrete stopping variable. Define

$$\mathbf{t}_n^* = \begin{cases} \frac{k}{n}, & \frac{k-1}{n} < \mathbf{t}^* \leq \frac{k}{n}, \quad k > 1, \\ \frac{1}{n}, & 0 \leq \mathbf{t}^* \leq \frac{1}{n}, \quad k = 1. \end{cases}$$

Then  $\mathbf{t}_n^*$  is a stopping time because for  $k/n \leq t < (k+1)/n$ ,

$$\{\mathbf{t}_n^* \leq t\} = \{\mathbf{t}^* \leq k/n\} \in \mathcal{F}(X(s), s \leq k/n).$$

But  $\mathcal{F}(X(s), s \leq k/n) \subset \mathcal{F}(X(s), s \leq t)$ . Also, I assert that

$$B \in \mathcal{F}(X(t), t \leq \mathbf{t}^*) \Rightarrow B \in \mathcal{F}(X(t), t \leq \mathbf{t}_n^*),$$

the latter because for  $k/n \leq t < (k+1)/n$ ,

$$B \cap \{\mathbf{t}_n^* \leq t\} = B \cap \{\mathbf{t}^* \leq k/n\} \in \mathcal{F}(X(s), s \leq t).$$

Let  $X_n^{(1)}(t) = X(t + \mathbf{t}_n^*) - X(\mathbf{t}_n^*)$ . By (12.43), for  $B \in \mathcal{F}(X(t), t \leq \mathbf{t}_n^*)$ ,

$$\begin{aligned} P(X_n^{(1)}(t_1) < x_1, \dots, X_n^{(1)}(t_j) < x_j, B) \\ = P(X(t_1) < x_1, \dots, X(t_j) < x_j) P(B). \end{aligned}$$

But, by the path continuity of  $X(t)$ ,  $X_n^{(1)}(t) \rightarrow X^{(1)}(t)$  for every  $\omega, t$ . This implies that at every point  $(x_1, \dots, x_j)$  which is a continuity point of the distribution function of  $X^{(1)}(t_1), \dots, X^{(1)}(t_j)$ ,

$$\begin{aligned} P(X^{(1)}(t_1) < x_1, \dots, X^{(1)}(t_n) < x_n, B) \\ = P(X(t_1) < x_1, \dots, X(t_n) < x_n) P(B). \end{aligned}$$

This is enough to ensure that equality holds for all  $(x_1, \dots, x_j)$ . Extension now proves the theorem.

**Problem 10.** Prove that the variables

$$\tau_1^* = \inf \{t; X(t) = 0, t \geq a\},$$

$$\tau_2^* = \inf \{t; X(t) = b\},$$

are stopping times for Brownian motion.

## NOTES

The motion of small particles suspended in water was noticed and described by Brown in 1828. The mathematical formulation and study was initiated by Bachelier [1, 1900], and Einstein [45, 1905], and carried on extensively from that time by both physicists and mathematicians. But rigorous discussion of sample path properties was not started until 1923 when Wiener [141] proved path continuity. Wiener also deduced the orthogonal expansion (12.24) in 1924 [142].

A good source for many of the deeper properties of Brownian motion is Lévy's books [103, 105], and in the recent book [76] by Ito and McKean. A very interesting collection of articles that includes many references to earlier works and gives a number of different ways of looking at Brownian motion has been compiled by Wax [139]. The article by Dvoretzki, Erdős, and Katutani [42] gives the further puzzling property that no sample paths have any "points of increase."

The fact of nondifferentiability of the sample paths was discovered by Paley, Wiener, and Zygmund [115, 1933]. The law of the iterated logarithm for Brownian motion was proved by Khintchine [88, 1933]. The properties of the zero sets of its paths was stated by Lévy, who seemed to assume the truth of 12.42. This latter property was stated and proved by Hunt [74, 1956]. David Freedman suggested the proof given that no zeros of  $X(t)$  are isolated.

## CHAPTER 13

# INVARIANCE THEOREMS

### 1. INTRODUCTION

Let  $S_1, S_2, \dots$  be a player's total winnings in a fair coin-tossing game. A question leading to the famous arc sine theorem is: Let  $N_n$  be the number of times that the player is ahead in the first  $n$  games,

$$N_n = \{\text{number of } k; k \leq n, S_k > 0\}.$$

The proportion of the time that the player is ahead in  $n$  games is  $N_n/n = W_n$ . Does a limiting distribution exist for  $W_n$ , and if so, what is it?

Reason this way: Define  $Z^{(n)}(t) = S_{[nt]}$  as  $t$  ranges over the values  $0 \leq t \leq 1$ . Denote Lebesgue measure by  $l$ ; then

$$W_n = l\{t; Z^{(n)}(t) > 0, 0 \leq t \leq 1\}.$$

Now  $Z^{(n)}(t) = S_{[nt]}$  does not converge to anything in any sense, but recall from Section 2 of the last chapter that the processes  $X^{(n)}(t) = S_{[nt]}/\sqrt{n}$  have all finite dimensional distribution functions converging to those of normalized Brownian motion  $X(t)$  as  $n \rightarrow \infty$ . We denote this by  $X^{(n)}(\cdot) \xrightarrow{\mathcal{D}} X(\cdot)$ . But  $W_n$  can also be written as

$$W_n = l\{t; X^{(n)}(t) > 0, 0 \leq t \leq 1\}.$$

Of course, the big transition that we would like to make here would be to define

$$W = l\{t; X(t) > 0, 0 \leq t \leq 1\},$$

so that  $W$  is just the proportion of time that a Brownian particle stays in the positive axis during  $[0, 1]$ , and then conclude that

$$W_n \xrightarrow{\mathcal{D}} W.$$

The general truth of an assertion like this would be a profound generalization of the central limit theorem. The transition from the obvious application of the central limit theorem to conclude that  $X^{(n)}(\cdot) \xrightarrow{\mathcal{D}} X(\cdot)$  to get to

$W_n \xrightarrow{\mathcal{D}} W$  is neither obvious or easy. But some general theorems of this kind would give an enormous number of limit theorems. For example, if we let

$$M_n = \max(|S_1|, \dots, |S_n|),$$

is it possible to find constants  $A_n$  such that  $M_n/A_n \xrightarrow{\mathcal{D}} M$ ,  $M$  nondegenerate? Again, write

$$M_n = \max_{t \in [0,1]} |S_{[nt]}| = \sqrt{n} \max_{t \in [0,1]} |X^{(n)}(t)|.$$

Let

$$M = \max_{t \in [0,1]} |X(t)|,$$

then apply our nonexistent theorem to conclude

$$M_n/\sqrt{n} \xrightarrow{\mathcal{D}} M.$$

This chapter will fill in the missing theorem. The core of the theorem is that successive sums of independent, identically distributed random variables with zero means and finite variances have the same distribution as Brownian motion with a random time index. The idea is not difficult—suppose  $S_1, S_2, \dots$  form the symmetric random walk. Define  $T_1$  as the first time such that  $|X(t)| = 1$ . By symmetry  $P(X(T_1) = 1) = P(X(T_1) = -1) = \frac{1}{2}$ . Define  $T_2$  as the first time such that  $|X(t + T_1) - X(T_1)| = 1$ , and so on. But  $T_1$  is determined only by the behavior of the  $X(t)$  motion up to time  $T_1$ , so that intuitively one might hope that the  $X(t + T_1) - X(T_1)$  process would have the same distribution as Brownian motion, but be independent of  $X(t)$ ,  $t \leq T_1$ . To make this sort of construction hold, the strong Markov property is, of course, essential. But also, we need to know some more about the first time that a Brownian motion exits from some interval around the origin.

## 2. THE FIRST-EXIT DISTRIBUTION

For some set  $B \in \mathcal{B}_1$ , let

$$\tau^*(B) = \inf \{t; X(t) \in B^c\}$$

be the first exit time of the Brownian motion from the set  $B$ . In particular, the first time that the particle hits the point  $\{a\}$  is identical with the first exit time of the particle from  $(-\infty, a)$  if  $a > 0$ , or from  $(\infty, a)$  if  $a < 0$ , and we denote it by  $\tau_a^*$ . Let  $\tau^*(a, b)$  be the first exit time from the interval  $(a, b)$ . The information we want is the probability that the first exit of the particle from the interval  $(a, b)$ ,  $a < 0 < b$ , occurs at the point  $b$ , and the value of  $E\tau^*(a, b)$ , the expected time until exit.

For normalized Brownian motion  $\{X(t)\}$ , use the substitution  $X(t) = X(s) + (X(t) - X(s))$  to verify that for  $s \leq t$ ,

$$E(X(t) \mid X(\tau), \tau \leq s) = X(s) \quad \text{a.s.}$$

$$E(X^2(t) - t \mid X(\tau), \tau \leq s) = X^2(s) - s \quad \text{a.s.}$$

Hence the processes  $\{X(t)\}$ ,  $\{X^2(t) - t\}$  are martingales in the sense of

**Definition 13.1.** A process  $\{Y(t)\}$ ,  $t \in I$ , is a martingale if for all  $t$  in  $I$ ,  $E|Y(t)| < \infty$ , and for any  $s, t$  in  $I$ ,  $s \leq t$ ,

$$E(Y(t) \mid Y(\tau), \tau \leq s) = Y(s) \quad \text{a.s.}$$

Suppose that for stopping times  $t^*$  satisfying the appropriate integrability conditions, the generalization of 5.31 holds as the statement  $EY(t^*) = EY(0)$ . This would give, for a stopping time on Brownian motion,

$$(13.2) \quad EX(t^*) = 0, \quad E(X^2(t^*) - t^*) = 0.$$

These two equations would give the parameters we want. If we take  $t^* = t^*(a, b)$ , (13.2) becomes

$$aP(X(t^*) = a) + bP(X(t^*) = b) = 0.$$

Solving, we get

$$(13.3) \quad P(X(t^*) = b) = \frac{|a|}{|b| + |a|}, \quad P(X(t^*) = a) = \frac{|b|}{|b| + |a|}.$$

The second equation of (13.2) provides us with

$$a^2P(X(t^*) = a) + b^2P(X(t^*) = b) = Et^*(a, b).$$

Using (13.3) we find that

$$(13.4) \quad Et^*(a, b) = |a| |b|.$$

Rather than detour to prove the general martingale result, we defer proof until the next chapter and prove here only what we need.

**Proposition 13.5.** For  $t^* = t^*(a, b)$ ,  $a < 0 < b$ ,

$$EX(t^*) = 0, \quad EX^2(t^*) = Et^*.$$

*Proof.* Let  $t^*$  be a stopping time taking values in a countable set  $\{t_j\} \subset [0, \tau]$ ,  $\tau < \infty$ . Then

$$E(X(\tau) - X(t^*)) = \sum_j \int_{\{t^* = t_j\}} (X(\tau) - X(t_j)) dP.$$

By independence, then,  $E(X(\tau) - X(t^*)) = 0$ ; hence  $EX(t^*) = 0$ . For the second equation, write

$$\begin{aligned} EX(\tau)^2 &= E(X(\tau) - X(t^*) + X(t^*))^2 \\ &= E(X(\tau) - X(t^*))^2 + EX^2(t^*) + 2E(X(\tau) - X(t^*))(X(t^*)). \end{aligned}$$

The last term is easily seen to be zero. The first term is

$$\sum_j \int_{\{t^* = t_j\}} (X(\tau) - X(t_j))^2 dP$$

which equals

$$\sum_j (\tau - t_j) P(t^* = t_j) = \tau - Et^*.$$

Since  $EX(\tau)^2 = \tau$ ,

$$EX^2(t^*) = Et^*.$$

Take  $t^{**} = \min(t^*(a, b), \tau)$ , and  $t_n^*$  a sequence of stopping times taking values in a countable subset of  $[0, \tau]$  such that  $t_n^* \rightarrow t^{**}$  everywhere. By the bounded convergence theorem,  $Et_n^* \rightarrow Et^{**}$ . Furthermore, by path continuity  $X(t_n^*) \rightarrow X(t^{**})$ , and

$$|X(t_n^*)|^2 \leq \sup_{t \leq \tau} |X(t)|^2,$$

which is integrable by 12.20. Use the bounded convergence theorem again to get  $EX(t_n^*) \rightarrow EX(t^{**})$ ,  $EX^2(t_n^*) \rightarrow EX^2(t^{**})$ . Write  $t^*$  for  $t^*(a, b)$ , then

$$\begin{aligned} \int X(t^{**}) dP &= \int_{t^* \leq \tau} X(t^*) dP + \int_{t^* > \tau} X(\tau) dP, \\ \int X^2(t^{**}) dP &= \int_{t^* \leq \tau} X^2(t^*) dP + \int_{t^* > \tau} X^2(\tau) dP. \end{aligned}$$

Note that  $|X(t^*)| \leq \max(|a|, |b|)$ , and on  $t^* > \tau$ ,  $|X(\tau)| \leq \max(|a|, |b|)$ . Hence as  $\tau \rightarrow \infty$ ,

$$EX(t^{**}) \rightarrow EX(t^*), \quad EX^2(t^{**}) \rightarrow EX^2(t^*).$$

Since  $t^{**} \uparrow t^*$ , apply monotone convergence to get  $\lim Et^{**} = Et^*$ , completing the proof.

**Problem 1.** Use Wald's identity 5.34 on the sums

$$S_n = \sum_{k=1}^n (X(k\Delta t) - X((k-1)\Delta t)),$$

and, letting  $\Delta t \rightarrow 0$ , prove that for  $t^* = t^*(a, b)$ , and any  $\lambda$ ,

$$E \exp [\lambda X(t^*) - \frac{1}{2} \lambda^2 t^*] = 1.$$

By showing that differentiation under the integral is permissible, prove 13.5.



## 3. REPRESENTATION OF SUMS

An important representation is given by

**Theorem 13.6.** *Given independent, identically distributed random variables  $Y_1, Y_2, \dots$ ,  $EY_1 = 0$ ,  $EY_1^2 = \sigma^2 < \infty$ ,  $S_n = Y_1 + \dots + Y_n$ , there exists a probability space with a Brownian motion  $X(t)$  and a sequence  $T_1, T_2, \dots$  of nonnegative, independent, identically distributed random variables defined on it such that the sequence  $X(T_1), X(T_1 + T_2), \dots$ , has the same distribution as  $S_1, S_2, \dots$ , and  $ET_1 = \sigma^2$ .*

*Proof.* Let  $(U_n, V_n)$ ,  $n = 1, 2, \dots$  be a sequence of identically distributed, independent random vectors defined on the same probability space as a Brownian motion  $\{X(t)\}$  such that  $\mathcal{F}(X(t), t \geq 0)$  and  $\mathcal{F}(U_n, V_n, n = 1, 2, \dots)$  are independent. This can be done by constructing  $(\Omega_1, \mathcal{F}_1, P_1)$  for the Brownian motion,  $(\Omega_2, \mathcal{F}_2, P_2)$  for the  $(U_n, V_n)$  sequence, and taking

$$(\Omega, \mathcal{F}, P) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, P_1 \times P_2).$$

Suppose  $U_n \leq 0 \leq V_n$ . Define

$$T_1 = \inf \{t; X(t) \in (U_1, V_1)^c\}.$$

Therefore the  $U_1, V_1$  function as random boundaries. Note that  $T_1$  is a random variable because

$$\{T_1 \leq t\} = \left\{ \inf_{0 \leq \tau \leq t} X(\tau) \leq U_1 \right\} \cup \left\{ \sup_{0 \leq \tau \leq t} X(\tau) \geq V_1 \right\}.$$

Further,  $\mathcal{F}(X(t + \tau) - X(\tau), t \geq 0)$  is independent of  $\mathcal{F}(X(s), s \leq \tau, U_1, V_1)$ . By the same argument which was used to establish the strong Markov property,

$$X^{(1)}(t) = X(t + T_1) - X(T_1), \quad t \geq 0$$

is a Brownian motion and is independent of  $X(T_1)$ . Now define  $T_2$  as the first exit time of  $X^{(1)}(t)$  from  $(U_2, V_2)$ . Then  $X(T_1 + T_2) - X(T_1)$  has the same distribution as  $X(T_1)$ . Repeating this procedure we manufacture variables

$$X(T_1 + \dots + T_k) - X(T_1 + \dots + T_{k-1}),$$

which are independent and identically distributed. The trick is to select  $U_1, V_1$  so that  $X(T_1)$  has the same distribution as  $Y_1$ . For any random boundaries  $U_1, V_1$ , if  $E|X(T_1)|^2 < \infty$ , 13.5 gives

$$E(X(T_1) | U_1, V_1) = 0 \quad \text{a.s.},$$

$$E(X^2(T_1) | U_1, V_1) = E(T_1 | U_1, V_1) \quad \text{a.s.}$$

Hence

$$EX(T_1) = 0, \quad EX^2(T_1) = ET_1.$$

Therefore if  $X(T_1)$  has the same distribution as  $Y_1$ , then  $Y_1$  must satisfy

$$E_1 Y = 0,$$

and automatically,

$$\sigma^2 = EY_1^2 = ET_1.$$

So  $EY_1 = 0$  is certainly a necessary condition for the existence of random boundaries such that  $X(T_1)$  has the same distribution as  $Y_1$ .

To show that it is also sufficient, start with the observation that if  $Y_1$  takes on only two values, say  $u < 0$ , and  $v > 0$  with probabilities  $p$  and  $q$ , then from  $EY_1 = 0$ , we have  $pu + qv = 0$ . For this distribution, we can take fixed boundaries  $U_1 = u$ ,  $V_1 = v$ . Because, by 13.5,  $EX(T_1) = 0$ ,  $uP(X(T_1) = u) + vP(X(T_1) = v) = 0$ , which implies that  $P(X(T_1) = u) = p$ . This idea can be extended to prove

**Proposition 13.7.** *For any random variable  $Y$ , such that  $EY = 0$ ,  $EY^2 < \infty$  there are random boundaries  $U \leq 0 \leq V$  such that  $X(T) \stackrel{D}{=} Y$ .*

*Proof.* Assume first that the distribution of  $Y$  is concentrated on points  $u_i \leq 0$ ,  $v_i \geq 0$  with probability  $p_i, q_i$  such that  $(u_i, v_i)$  are pairs satisfying  $u_i p_i + v_i q_i = 0$ . Then take

$$P((U, V) = (u_i, v_i)) = p_i + q_i.$$

By the observation above,

$$P(X(T) = u_i | (U, V) = (u_i, v_i)) = \frac{p_i}{p_i + q_i}.$$

Therefore  $X(T)$  has the same distribution as  $Y$ . Suppose now that the distribution of  $Y$  is concentrated on a finite set  $\{y_j\}$  of points. Then it is easy to see that the pairs  $u_i \leq 0$ ,  $v_i \geq 0$  can be gotten such that  $Y$  assumes only values in  $\{u_i\}$ ,  $\{v_i\}$  and the pairs  $(u_i, v_i)$  satisfy the conditions above. (Note that  $u_i$  may equal  $u_j$ ,  $i \neq j$ , and similarly the  $v_i$  are not necessarily distinct.) For  $Y$  having any distribution such that  $EY = 0$ ,  $EY^2 < \infty$ , take  $Y_n \xrightarrow{D} Y$  where  $Y_n$  takes values in a finite set of points,  $EY_n = 0$ . Define random boundaries  $U_n, V_n$  having stopping time  $T_n$  such that  $X(T_n) \stackrel{D}{=} Y_n$ . Suppose that the random vectors  $(U_n, V_n)$  have mass-preserving distributions. Then take  $(U_n, V_n) \xrightarrow{D} (U, V)$ . For these random boundaries and associated stopping time  $T$ , for  $I \subset (0, \infty)$ , use (13.3) to get

$$P(X(T) \in I | U, V) = \chi_I(V) \frac{|U|}{|U| + |V|} = g(U, V).$$

Similarly,

$$P(X(T_n) \in I | U_n, V_n) = g(U_n, V_n).$$

Hence if  $P(Y \in bd(I)) = P(V \in bd(I)) = 0$ , then

$$\begin{aligned} P(Y \in I) &= \lim_n P(Y_n \in I) \\ &= \lim_n Eg(U_n, V_n) \\ &= Eg(U, V) \\ &= P(X(T) \in I). \end{aligned}$$

The analogous proof holds for  $I \subset (0, -\infty)$ . To complete the proof we need therefore to show that the  $Y_n$  can be selected so that the  $(U_n, V_n)$  have a mass-preserving set of distributions. Take  $F(dy)$  to denote the distribution of  $Y$ . We can always select a nonempty finite interval  $(a, b)$  such that including part or all of the mass of  $F(dy)$  at the endpoints of  $(a, b)$  in the integral we get

$$\int_a^b yF(dy) = 0.$$

Thus we can always take the distributions  $F_n$  of the  $Y_n$  such that

$$\int_{[a,b]} yF_n(dy) = 0.$$

In this case, the  $(u_i, v_i)$  pairs have the property that either both are in  $[a, b]$  or both are outside of  $[a, b]$ . Since  $EY^2 < \infty$ , we can also certainly take the  $Y_n$  such that  $EY_n^2 \leq M < \infty$  for all  $n$ . Write

$$\begin{aligned} EY_n^2 &= E(E(X^2(T_n) \mid U_n, V_n)) = E(E(T_n \mid U_n, V_n)) \\ &= E|U_n V_n|. \end{aligned}$$

But the function  $|uv|$  goes to infinity as either  $|u| \rightarrow \infty$  or  $|v| \rightarrow \infty$  everywhere in the region  $\{u < a, v > b\}$ . This does it.

**Problem 2.** Let the distribution function of  $Y$  be  $F(dy)$ . Prove that the random boundaries  $(U, V)$  with distribution

$$\begin{aligned} P(U \in du) &= F(du) \\ P(V \in dv \mid U = u) &= \alpha |v| g(u, v) F(dv), \end{aligned}$$

where  $\alpha^{-1} = EY^+$  and  $g(u, v)$  is zero or one as  $u$  and  $v$  have the same or opposite signs, give rise to an exit time  $T$  such that  $X(T) \stackrel{D}{=} Y$ . (Here  $U$  and  $V$  can be both positive and negative.)

#### 4. CONVERGENCE OF SAMPLE PATHS OF SUMS TO BROWNIAN MOTION PATHS

Now it is possible to show that in a very strong sense Brownian motion is the limit of random walks with smaller and smaller steps, or of normed sums of independent, identically distributed random variables. The random

walk example is particularly illuminating. Let the walk  $X^{(n)}(t)$  take steps of size  $\pm 1/\sqrt{n}$  every  $1/n$  time units. Using the method of the previous section, let  $T_1$  be the first time that  $X(t)$  changes by an amount  $1/\sqrt{n}$ , then  $T_2$  the time until a second change of amount  $1/\sqrt{n}$  occurs, etc.

The process

$$\tilde{X}^{(n)}(t) = X(T_1 + \cdots + T_{[nt]})$$

has the same distribution as the process  $X^{(n)}(t)$ . By definition,

$$T_1 + \cdots + T_{[nt]} = \text{time until } [nt] \text{ changes of magnitude } 1/\sqrt{n} \text{ have occurred along } X(t).$$

Therefore, up to time  $T_1 + \cdots + T_{[nt]}$ , the sum of the squares of the changes in  $X(t)$  is approximately  $t$ . But by 12.28, this takes a length of time  $t$ . So, we would expect that  $T_1 + \cdots + T_{[nt]} \rightarrow t$ , hence that each sample path of the interpolated motion  $\tilde{X}^{(n)}(t)$  would converge as a function of  $t$  to the corresponding path of Brownian motion. The convergence that does take place is uniform convergence. This holds, in general, along subsequences.

**Theorem 13.8.** Let  $Y_1, Y_2, \dots$  be independent, identically distributed random variables,  $EY_1 = 0$ ,  $EY_1^2 = \sigma^2 < \infty$ ,  $S_n = Y_1 + \cdots + Y_n$ . Define the processes  $X^{(n)}(t)$  by  $S_{[nt]}/\sigma\sqrt{n}$ . Then there are processes  $\{\tilde{X}^{(n)}(t)\}$ , for each  $n$  having the same distribution as  $\{X^{(n)}(t)\}$ , defined on a common probability space and a Brownian motion process  $\{X(t)\}$  on the same space, such that for any subsequence  $\{n_k\}$  increasing rapidly enough,

$$\sup_{0 \leq t \leq 1} |\tilde{X}^{(n_k)}(t) - X(t)| \rightarrow 0 \quad \text{a.s.}$$

*Proof.* Assume that  $EY_1^2 = 1$ . Let  $(\Omega, \mathcal{F}, P)$  be constructed as in the representation theorem. For each  $n$ , consider the Brownian motion  $X_n(t) = \sqrt{n}X(t/n)$ . Construct  $T_1^{(n)}, T_2^{(n)}, \dots$  using the motion  $X_n(t)$ . Then the  $\{S_k\}$  sequence has the same distribution as the  $\{X_n(T_1^{(n)} + \cdots + T_k^{(n)})\}$  sequence. Thus the  $X^{(n)}(t)$  process has the same distribution as

$$(13.9) \quad \tilde{X}^{(n)}(t) = X\left(\frac{T_1^{(n)} + \cdots + T_{[nt]}^{(n)}}{n}\right).$$

The sequence  $T_1^{(n)}, T_2^{(n)}, \dots$  for each  $n$  consists of independent, identically distributed random variables such that  $ET_1^{(n)} = 1$ , and  $T_1^{(n)}$  has the same distribution for all  $n$ . The weak law of large numbers gives

$$\frac{T_1^{(n)} + \cdots + T_{[nt]}^{(n)}}{[nt]} \xrightarrow{P} 1,$$

so for  $n$  large, the random time appearing in (13.9) should be nearly  $t$ .

Argue that if it can be shown that

$$W_n = \sup_{0 \leq t \leq 1} \left| \frac{T_1^{(n)} + \cdots + T_{[nt]}^{(n)}}{n} - t \right| \rightarrow 0 \quad \text{a.s.}$$

for  $n$  running through some subsequence, then the continuity of  $X(t)$  guarantees that along the same subsequence

$$\sup_{0 \leq t \leq 1} |X^{(n)}(t) - X(t)| \rightarrow 0 \quad \text{a.s.}$$

What can be easily proved is that  $W_n \xrightarrow{P} 0$ . But this is enough because then for any subsequence increasing rapidly enough,  $W_{n_k} \xrightarrow{\text{a.s.}} 0$ . Use

$$\bar{T}_k^{(n)} = T_k^{(n)} - 1,$$

so

$$E\bar{T}_k^{(n)} = 0,$$

$$W_n = \sup_{0 \leq t \leq 1} \left| \frac{\bar{T}_1^{(n)} + \cdots + \bar{T}_{[nt]}^{(n)}}{n} \right| + \frac{\theta}{n}, \quad |\theta| \leq 1.$$

Ignore the second term, and write

$$W_n \leq \sup_{0 \leq t \leq 1} t \left| \frac{\bar{T}_1^{(n)} + \cdots + \bar{T}_{[nt]}^{(n)}}{[nt]} \right|.$$

For any  $\epsilon$ ,  $0 < \epsilon < 1$ , take

$$\begin{aligned} W'_n &= \epsilon \sup_{0 \leq t \leq \epsilon} \left| \frac{\bar{T}_1^{(n)} + \cdots + \bar{T}_{[nt]}^{(n)}}{[nt]} \right| \\ &\leq \epsilon \sup_{k \geq 1} \left| \frac{\bar{T}_1^{(n)} + \cdots + \bar{T}_k^{(n)}}{k} \right| = \epsilon M^{(n)}. \end{aligned}$$

The distribution of  $M^{(n)}$  is the same as  $M^{(1)}$ . Now write

$$\begin{aligned} W''_n &= \sup_{\epsilon \leq t \leq 1} \left| \frac{\bar{T}_1^{(n)} + \cdots + \bar{T}_{[nt]}^{(n)}}{[nt]} \right| \\ &\leq \sup_{k \geq [\epsilon n]} \left| \frac{\bar{T}_1^{(n)} + \cdots + \bar{T}_k^{(n)}}{k} \right|. \end{aligned}$$

This bounding term has the same distribution as

$$\sup_{k \geq [\epsilon n]} \left| \frac{\bar{T}_1^{(1)} + \cdots + \bar{T}_k^{(1)}}{k} \right|.$$

The law of large numbers implies

$$\frac{\bar{T}_1^{(1)} + \cdots + \bar{T}_k^{(1)}}{k} \xrightarrow{\text{a.s.}} 0,$$

so that  $W_n'' \xrightarrow{P} 0$ . Since  $W_n \leq W_n' + W_n''$ , then for any  $\epsilon > 0$ ,  $x > 0$ ,

$$\overline{\lim} P(W_n > x) \leq P(\epsilon M^{(1)} > x).$$

Taking  $\epsilon \downarrow 0$  now gives the result.

## 5. AN INVARIANCE PRINCIPLE

The question raised in the introduction to this chapter is generally this:

*The sequence of processes  $X^{(n)}(\cdot)$  converges in distribution to Brownian motion  $X(\cdot)$ , denoted  $X^{(n)}(\cdot) \xrightarrow{\mathcal{D}} X(\cdot)$ , in the sense that for any  $0 \leq t_1 < \cdots < t_k \leq 1$ ,*

$$(X^{(n)}(t_1), \dots, X^{(n)}(t_k)) \xrightarrow{\mathcal{D}} (X(t_1), \dots, X(t_k)).$$

*Let  $H(x(\cdot))$  be defined on  $R^{[0,1]}$ . When is it true that*

$$H(X^{(n)}(\cdot, \omega)) \xrightarrow{\mathcal{D}} H(X(\cdot, \omega))?$$

There is no obvious handle. What is clear, however, is that we can proceed as follows: Let us suppose  $X^{(n)}(\cdot)$  and  $X(\cdot)$  are defined on the same space and take values in some subset  $D \subset R^{[0,1]}$ . On  $D$  define the sup-norm metric

$$\rho(x(\cdot), y(\cdot)) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|,$$

and assume that  $H(x(\cdot))$  defined on  $D$  is continuous with respect to  $\rho$ . Then if the sample paths of the  $X^{(n)}(\cdot)$  processes converge uniformly to the corresponding paths of  $X(\cdot)$ , that is, if

$$\sup_{0 \leq t \leq 1} |X^{(n)}(t) - X(t)| \rightarrow 0 \quad \text{a.s.},$$

then

$$H(X^{(n)}(\cdot)) \xrightarrow{\text{a.s.}} H(X(\cdot)).$$

But this is enough to give us what we want. Starting with the  $X^{(n)}(t) = S_{[nt]}/\sigma\sqrt{n}$ , we can construct  $\tilde{X}^{(n)}(\cdot)$  having the same distribution as  $X^{(n)}(\cdot)$  so that the  $\tilde{X}^{(n)}(t) \rightarrow X(t)$  uniformly for  $t \in [0, 1]$  for  $n$  running through subsequences. Thus,  $H(\tilde{X}^{(n)}(\cdot)) \xrightarrow{\text{a.s.}} H(X(\cdot))$ ,  $n \in \{n_k\}$ . This implies

$$H(X^{(n)}(\cdot)) \xrightarrow{\mathcal{D}} H(X(\cdot)), \quad n \in \{n_k\}.$$

But this holding true for every subsequence  $\{n_k\}$  increasing rapidly enough implies that the full sequence converges in distribution to  $H(X(\cdot))$ . Now to fasten down this idea.

**Definition 13.10.**  $D$  is the class of all functions  $x(t)$ ,  $0 \leq t \leq 1$ , such that  $x(t-)$ ,  $x(t+)$  exist for all  $t \in (0, 1)$ , and  $x(t+) = x(t)$ . Also,  $x(0+) = x(0)$ ,  $x(1-) = x(1)$ . Define  $\rho(x(\cdot), y(\cdot))$  on  $D$  by

$$\sup_{0 \leq t \leq 1} |x(t) - y(t)|.$$

**Definition 13.11.** For  $H(x(\cdot))$  defined on  $D$ , let  $G$  be the set of all functions  $x(\cdot) \in D$  such that  $H$  is discontinuous at  $x(\cdot)$  in the metric  $\rho$ . If there is a set  $G_1 \in \mathcal{B}^{[0,1]}$  such that  $G \subset G_1$  and for a normalized Brownian motion  $\{X(t)\}$ ,  $P(X(\cdot) \in G_1) = 0$ , call  $H$  a.s.  $B$ -continuous.

The weakening of the continuity condition on  $H$  to a.s.  $B$ -continuity is important. For example, the  $H$  that leads to the arc sine law is discontinuous at the set of all  $x(\cdot) \in D$  such that

$$l\{t; x(t) = 0, \quad 0 \leq t \leq 1\} > 0.$$

(We leave this to the reader to prove as Problem 4.) But this set has probability zero in Brownian motion.

With these definitions, we can state the following special case of the "invariance principle."

**Theorem 13.12.** Let  $H$  defined on  $D$  be a.s.  $B$ -continuous. Consider any process of the type

$$X^{(n)}(t) = \frac{S_{[nt]}}{\sigma\sqrt{n}},$$

where the  $S_n$  are sums of independent, identically distributed random variables  $Y_1, Y_2, \dots$  with  $EY_1 = 0$ ,  $EY_1^2 = \sigma^2$ . Assume that the  $H(X^{(n)}(\cdot))$  are random variables. Then

$$(13.13) \quad H(X^{(n)}(\cdot)) \xrightarrow{\mathcal{D}} H(X(\cdot)),$$

where  $\{X(t)\}$  is normalized Brownian motion.

*Proof.* Use 8.8; it is enough to show that any subsequence  $\{n_k\}$  contains a subsequence  $\{n'_k\}$  such that (13.13) holds along  $n'_k$ . Construct  $\tilde{X}_{n'_k}(t)$  as in the proof of 13.8. Take  $n'_k$  any subsequence of  $n_k$  increasing rapidly enough. Then  $\tilde{X}_{(n'_k)}(t)$  converges uniformly to  $X(t)$  for almost every  $\omega$ , implying that (13.13) holds along the  $n'_k$  sequence.

There is a loose end in that  $H(X(\cdot))$  was not assumed to be a random variable. However, since

$$H(\tilde{X}_{(n'_k)}(\cdot)) \xrightarrow{\text{a.s.}} H(X(\cdot)),$$

the latter is a.s. equal to a random variable. Hence it is a random variable

with respect to the completed probability space  $(\Omega, \mathcal{F}, \bar{P})$ , and its distribution is well defined.

The reason that theorems of this type are referred to as *invariance principles* is that they establish convergence to a limiting distribution which does not depend on the distribution function of the independent summands  $Y_1, Y_2, \dots$  except for the one parameter  $\sigma^2$ . This gives the freedom to choose the most convenient way to evaluate the limit distribution. Usually, this is done either directly for Brownian motion or by combinatorial arguments for coin-tossing variables  $Y_1, Y_2, \dots$ . In particular, see Feller's book [59, Vol. I], for a combinatorial proof that in fair coin-tossing, the proportion of times  $W_n$  that the player is ahead in the first  $n$  tosses has the limit distribution

$$\lim P(W_n < x) = 2/\pi \arcsin \sqrt{x}.$$

### Problems

3. Show that the function on  $D$  defined by

$$H(x(\cdot)) = \sup_{0 \leq t \leq 1} |x(t)|$$

is continuous everywhere.

4. Show that the function on  $D$  defined by

$$H(x(\cdot)) = I\{t; x(t) > 0, \quad 0 \leq t \leq 1\}$$

is continuous at  $x(\cdot)$  if and only if  $I\{t; x(t) = 0\} = 0$ .

## 6. THE KOLMOGOROV-SMIRNOV STATISTICS

An important application of invariance is to an estimation problem. Let  $Y_1, Y_2, \dots$  be independent, identically distributed random variables with a continuous but unknown distribution function  $F(x)$ . The most obvious way to estimate  $F(x)$  given  $n$  observations  $Y_1, \dots, Y_n$  is to put

$$\begin{aligned} \hat{F}_n(x) &= \frac{1}{n} (\text{number of } Y_k < x, \quad k = 1, \dots, n) \\ &= \frac{1}{n} \sum_{k=1}^n \chi_{(-\infty, x)}(Y_k). \end{aligned}$$

The law of large numbers guarantees

$$\hat{F}_n(x) \xrightarrow{\text{a.s.}} F(x)$$

for fixed  $x$ . From the central limit theorem,

$$\sqrt{n} (\hat{F}_n(x) - F(x)) \xrightarrow{D} \mathcal{N}(0, F(x)(1 - F(x))).$$



However, we will be more interested in uniform estimates:

$$D_n^+ = \sqrt{n} \sup_x (\hat{F}_n(x) - F(x)),$$

$$D_n^- = \sqrt{n} \inf_x (\hat{F}_n(x) - F(x)),$$

$$D_n = \sqrt{n} \sup_x |\hat{F}_n(x) - F(x)|,$$

and the problem is to show that  $D_n^+$ ,  $D_n^-$ ,  $D_n$  converge in distribution, and to find the limiting distribution.

**Proposition 13.14.** *Each of  $D_n^+$ ,  $D_n^-$ ,  $D_n$  has the same distribution for all continuous  $F(x)$ .*

*Proof.* Call  $I = [a, b]$  an interval of constancy for  $F(x)$  if  $P(Y_1 \in I) = 0$  and there is no larger interval containing  $I$  having this property. Let  $B$  be the union of all the intervals of constancy. Clearly, we can write

$$D_n = \sqrt{n} \sup_{x \in B^c} |\hat{F}_n(x) - F(x)|,$$

and similar equations for  $D_n^+$  and  $D_n^-$ . For  $x \in B^c$ , the sets

$$\{Y_k < x\}, \quad \{F(Y_k) < F(x)\}$$

are identical. Put  $U_k = F(Y_k)$ , and set

$$\hat{G}_n(y) = \{\text{number of } U_k < y, k = 1, \dots, n\}.$$

Then

$$\begin{aligned} D_n &= \sqrt{n} \sup_{x \in B^c} |\hat{G}_n(F(x)) - F(x)| \\ &= \sqrt{n} \sup_x |\hat{G}_n(F(x)) - F(x)|. \end{aligned}$$

Since  $F(x)$  maps  $R^{(1)}$  onto  $(0, 1)$  plus the points  $\{0\}$  or  $\{1\}$  possibly,

$$\begin{aligned} D_n &= \sqrt{n} \sup_{y \in (0,1)} |\hat{G}_n(y) - y| \\ &= \sqrt{n} \sup_{y \in [0,1]} |\hat{G}_n(y) - y| \quad \text{a.s.,} \end{aligned}$$

the latter holding because  $P(U_1 = 0) = P(U_1 = 1) = 0$ . The distribution of  $U_1$  is given by

$$P(U_1 < y) = P(F(Y_1) < y).$$

Put  $x = \inf \{\xi; F(\xi) = y\}$ . Then

$$P(U_1 < y) = P(F(Y_1) < F(x)) = P(Y_1 < x) = F(x) = y.$$

Thus  $U_1$  is uniformly distributed on  $[0, 1]$ , and  $D_n$  for arbitrary continuous

$F$  has the same distribution as  $D_n$  for the uniform distribution. Similarly for  $D_n^-$ , and  $D_n^+$ .

Let  $U_1, \dots, U_n$  be independent random variables uniformly distributed on  $[0, 1]$ . The order statistics are defined as follows:  $U_1^{(n)}$  is the smallest, and so forth;  $U_n^{(n)}$  is the largest. The maximum of  $|\hat{G}_n(y) - y|$  or of  $\hat{G}_n(y) - y$  or  $y - \hat{G}_n(y)$  must occur at one of the jumps of  $\hat{G}_n(y)$ . The jumps are at the points  $U_k^{(n)}$ , and

$$\hat{G}_n(U_k^{(n)}) = \frac{k-1}{n}.$$

Since the size of the jumps is  $1/n$ , then to within  $1/\sqrt{n}$ ,

$$\begin{aligned} D_n &= \sqrt{n} \max_{k \leq n} \left| U_k^{(n)} - \frac{k}{n} \right|, \\ D_n^+ &= \sqrt{n} \max_{k \leq n} \left( \frac{k}{n} - U_k^{(n)} \right), \\ D_n^- &= \sqrt{n} \min_{k \leq n} \left( \frac{k}{n} - U_k^{(n)} \right). \end{aligned}$$

The fact that makes our invariance theorem applicable is that the  $U_1^{(n)}, \dots, U_n^{(n)}$  behave something like sums of independent random variables. Let  $W_1, W_2, \dots$  be independent random variables with the negative exponential distribution. That is,

$$P(W_k > x) = e^{-x}, \quad x \geq 0.$$

Denote  $Z_n = W_1 + \dots + W_n$ ; then

**Proposition 13.15.**  $U_k^{(n)}$ ,  $k = 1, \dots, n$  have the same joint distribution as  $Z_k/Z_{n+1}$ ,  $k = 1, \dots, n$ .

*Proof.* To show this, write (using a little symbolic freedom),

$$\begin{aligned} P(Z_1 \in dx_1, Z_2 \in dx_2, \dots, Z_{n+1} \in dx_{n+1}) \\ &= P(W_1 \in dx_1, W_2 \in dx_2 - x_1, \dots, W_{n+1} \in dx_{n+1} - x_n) \\ &= e^{-x_1} e^{-(x_2-x_1)} \dots e^{-(x_{n+1}-x_n)} dx_1 \dots dx_{n+1} \\ &= e^{-x_{n+1}} dx_1 \dots dx_{n+1}, \quad 0 \leq x_1 \leq x_2 \leq \dots \leq x_{n+1}. \end{aligned}$$

Thus

$$\begin{aligned} P(Z_1 \in dx_1, \dots, Z_n \in dx_n \mid Z_{n+1} = x_{n+1}) \\ &= \begin{cases} n! x_{n+1}^{-n} dx_1 \dots dx_n, & 0 \leq x_1 \leq x_2 \leq \dots \leq x_{n+1}, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

From this,

$$P\left(\frac{Z_1}{Z_{n+1}} \in dy_1, \dots, \frac{Z_n}{Z_{n+1}} \in dy_n \mid Z_{n+1} = x_{n+1}\right) \\ = \begin{cases} n! dy_1 \cdots dy_n, & 0 \leq y_1 \leq \cdots \leq y_n \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$P\left(\frac{Z_1}{Z_{n+1}} \in dy_1, \dots, \frac{Z_n}{Z_{n+1}} \in dy_n\right) \\ = \begin{cases} n! dy_1 \cdots dy_n, & 0 \leq y_1 \leq \cdots \leq y_n \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

On the other hand, for  $0 \leq y_1 \leq \cdots \leq y_n \leq 1$ ,

$$P(U_1^{(n)} \in dy_1, \dots, U_n^{(n)} \in dy_n) = \sum P(U_{l_1} \in dy_1, \dots, U_{l_n} \in dy_n),$$

where the sum is over all permutations  $(l_1, \dots, l_n)$  of  $(1, \dots, n)$ . Using independence this yields  $n! dy_1 \cdots dy_n$ .

Use this proposition to transform the previous expression for  $D_n$  into

$$D_n \stackrel{\mathcal{D}}{=} \sqrt{n} \max_{k \leq n} \left| \frac{Z_k}{Z_{n+1}} - \frac{k}{n} \right|,$$

with analogous expressions for  $D_n^+$ ,  $D_n^-$ . Then

$$D_n \stackrel{\mathcal{D}}{=} \frac{n}{Z_{n+1}} \max_{k \leq n} \left| \frac{Z_k - k}{\sqrt{n}} - \frac{k}{n} \cdot \frac{Z_{n+1} - n}{\sqrt{n}} \right|.$$

Because  $EW_1 = 1$ ,  $\sigma^2(W_1) = 1$ , it follows that  $n/Z_{n+1} \xrightarrow{\text{a.s.}} 1$ , and that  $Z_k - k$  is a sum of independent, identically distributed random variables with first moment zero and second moment one. Put  $S_k = Z_k - k$ ,

$$X^{(n)}(t) = \frac{S_{[nt]}}{\sqrt{n}},$$

and ignore the  $n/Z_{n+1}$  term and terms of order  $1/\sqrt{n}$ . Then  $D_n$  is given by

$$\sup_{0 \leq t \leq 1} |X^{(n)}(t) - tX^{(n)}(1)|.$$

Obviously,

$$\sup_{0 \leq t \leq 1} |x(t) - tx(1)|$$

is a continuous function in the sup-norm metric, so now applying the invariance principle, we have proved

**Theorem 13.16**

$$D_n \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} |X(t) - tX(1)|,$$

$$D_n^+ \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} (X(t) - tX(1)),$$

$$D_n^- \xrightarrow{\mathcal{D}} \inf_{0 \leq t \leq 1} (X(t) - tX(1)).$$

**7. MORE ON FIRST-EXIT DISTRIBUTIONS**

There is a wealth of material in the literature on evaluating the distributions of functions on Brownian motion. One method uses some transformations that carry Brownian motion into Brownian motion. A partial list of such transformations is

**Proposition 13.17.** *If  $X(t)$  is normalized Brownian motion, then so is*

- 1)  $-X(t)$ ,  $t \geq 0$  (symmetry),
- 2)  $X(t + \tau) - X(\tau)$ ,  $t \geq 0$ ,  $\tau \geq 0$  fixed (origin change),
- 3)  $tX(1/t)$ ,  $t \geq 0$  (inversion),
- 4)  $(1/\sqrt{\alpha})X(\alpha t)$ ,  $t \geq 0$ ,  $\alpha > 0$  (scale change),
- 5)  $X(T) - X(T - t)$ ,  $0 \leq t \leq T$  fixed (reversal).

To get (4) and (5) just check that the processes are Gaussian with zero means and the right covariance.

We apply these transformations and the strong Markov property to get the distributions of some first exit times and probabilities. These are related to a number of important functions on Brownian motion. For example, for  $x > 0$ , if  $\tau_x^*$  is the first hitting time of the point  $\{x\}$ , then

$$(13.18) \quad P\left(\sup_{0 \leq \tau \leq t} X(\tau) < x\right) = P(\tau_x^* > t).$$

To get the distribution of  $\tau_x^*$ , let

$$\varphi(\lambda, x) = Ee^{-\lambda \tau_x^*}, \quad \lambda \geq 0.$$

Take  $x, y > 0$ , note that  $\tau_{x+y}^* = \tau_x^* + \tau_y^*$ , where  $\tau_y^*$  is the first passage time of the process  $X(t) = X(\tau_x^* + t) - X(\tau_x^*)$  to the point  $y$ . By the strong Markov property,  $\tau_y^*$  has the same distribution as  $\tau_y^*$  and is independent of  $\tau_x^*$ . Thus,

$$(13.19) \quad \varphi(\lambda, x + y) = \varphi(\lambda, x)\varphi(\lambda, y).$$

Since  $\varphi(\lambda, x)$  is decreasing in  $x$ , and therefore well-behaved, 13.19 implies

$$(13.20) \quad \varphi(\lambda, x) = e^{-\varphi(\lambda)x}.$$

Now we can get more information by a scale change. Transformation 13.17(4) implies that a scale change in space by an amount  $\sqrt{\alpha}$  changes time by a factor  $\alpha$ . To be exact,

$$\begin{aligned} t_x^* &= \inf \{t; X(t) = x\} \\ &= \alpha \inf \{t/\alpha; (1/\sqrt{\alpha})X(t/\alpha) = x/\sqrt{\alpha}\} \\ &= \alpha \inf \{t'; X(t') = x/\sqrt{\alpha}\}. \end{aligned}$$

Therefore  $t_x^*$  has the same distribution as  $\alpha t_{x/\sqrt{\alpha}}^*$ , yielding

$$\begin{aligned} (13.21) \quad \varphi(\lambda, x) &= \varphi(\alpha\lambda, x/\sqrt{\alpha}), \\ \Rightarrow g(\lambda) &= c\sqrt{\lambda}, \\ \Rightarrow \varphi(x, \lambda) &= e^{-c\sqrt{\lambda}x}, \quad x > 0, \\ \Rightarrow \varphi(x, \lambda) &= e^{-c\sqrt{\lambda}|x|}, \quad \text{all } x, \quad \text{by symmetry.} \end{aligned}$$

Now  $\varphi(x, \lambda)$  uniquely determines the distribution of  $t_x^*$ , so if we can get  $c$ , then we are finished. Unfortunately, there seems to be no very simple way to get  $c$ . Problem 5 outlines one method of showing that  $c = \sqrt{2}$ . Accept this for now, because arguments of this sort can get us the distribution of

$$t^*(a, b) = \min(t_a^*, t_b^*), \quad \{0\} \in (a, b).$$

Denote

$$t^* = \min(t_a^*, t_b^*), \quad A_a = \{t_a^* < t_b^*\}, \quad A_b = \{t_b^* < t_a^*\},$$

so that on  $A_a$ ,  $t_a^* = t^*$ , and on  $A_b$ ,  $t_a^* = t^* + \tau^*$ , where  $\tau^*$  is the additional time needed to get to  $x = a$  once the process has hit  $x = b$ . So define

$$\tau^* = \min \{t; X(t^* + t) - X(t^*) = a - b\}.$$

Put these together:

$$Ee^{-\lambda t_a^*} = \int_{A_a} e^{-\lambda t^*} dP + \int_{A_b} e^{-\lambda t^*} e^{-\lambda \tau^*} dP.$$

Now check that  $A_b \in \mathcal{F}(X(t), t \leq t^*)$ . Since the variable  $\tau^*$  is independent of  $\mathcal{F}(X(t), t \leq t^*)$ ,

$$e^{-\sqrt{2\lambda}|a|} = \int_{A_a} e^{-\lambda t^*} dP + e^{-\sqrt{2\lambda}(|a|+|b|)} \int_{A_b} e^{-\lambda t^*} dP.$$

The same argument for  $t_b^*$  gives

$$(13.22) \quad e^{-\sqrt{2\lambda}|b|} = e^{-\sqrt{2\lambda}(|b|+|a|)} \int_{A_a} e^{-\lambda t^*} dP + \int_{A_b} e^{-\lambda t^*} dP.$$

Now solve, to get

$$(13.23) \quad 1) \int_{\mathcal{A}_a} e^{-\lambda \tau^*} dP = \frac{\sinh(\sqrt{2\lambda} |b|)}{\sinh(\sqrt{2\lambda}(|b| + |a|))},$$

$$2) \int_{\mathcal{A}_b} e^{-\lambda \tau^*} dP = \frac{\sinh(\sqrt{2\lambda} |a|)}{\sinh(\sqrt{2\lambda}(|b| + |a|))}.$$

The sum of (13.23(1) and (2)) is  $Ee^{-\lambda \tau^*}$ , the Laplace transform of the distribution of  $\tau^*$ . By inverting this, we can get

$$P(a < X(\tau) < b, \text{ all } \tau \in [0, t]).$$

Very similar methods can be used to compute the probability that the Brownian motion ever hits the line  $x = at + b$ ,  $a \geq 0, b \geq 0$ , or equivalently, exits from the open region with the variable boundary  $x = at + b$ . Let  $p(a, b)$  be the probability that  $X(t)$  ever touches the line  $at + b$ ,  $a \geq 0, b \geq 0$ . Then

$$p(a, b_1 + b_2) = p(a, b_1)p(a, b_2),$$

the argument being that to get to  $at + b_1 + b_2$ , first the particle must get to  $at + b_1$ , but once it does, it then has to get to a line whose equation relative to its present position is  $at + b_2$ . To define this more rigorously, let  $\tau^*$  be the time of first touching  $at + b_1$ ; then  $\tau^* = \min(\tau^*, s)$  is a stopping time. The probability that the process ever touches the line  $at + b_1 + b_2$  and  $\tau^* < s$  equals the probability that the process  $X(t + \tau^*) - X(\tau^*)$  ever touches the line  $at + b_2$  and  $\tau^* < s$ . By the strong Markov property, the latter probability is the product  $p(a, b_2)P(\tau^* < s)$ . Let  $s \rightarrow \infty$  to get the result. Therefore,  $p(a, b) = e^{-\psi(a)b}$ . Take  $\tau_b^*$  to be the hitting time of the point  $b$ . Then

$$p(a, b) = P\left(\sup_{0 \leq t < \infty} \{X(t + \tau_b^*) - X(\tau_b^*) - (at + a\tau_b^*)\} \geq 0\right).$$

Conditioning on  $\tau_b^*$  yields

$$p(a, b) = Ee^{-\psi(a)a\tau_b^*},$$

(see 4.38). Use (13.21) to conclude that

$$p(a, b) = e^{-\sqrt{2a\psi(a)}b},$$

which leads to  $2a\psi(a) = \psi^2(a)$ , or  $\psi(a) = 2a$ . Thus

$$(13.24) \quad p(a, b) = e^{-2ab}.$$

The probability,

$$P\left(\sup_t \frac{|X(t)|}{at + b} \geq 1\right),$$

of exiting from the two-sided region  $|x| < at + b$  is more difficult to compute. One way is to first compute  $Ee^{-\lambda t^*}$ , where  $t^*$  is the first time of hitting  $at + b$ , and then imitate the development leading to the two-sided boundary distribution in (13.23). Another method is given in Doob [36].

The expression for  $p(a, b)$  can be used to get the distribution of

$$\sup_{0 \leq t \leq 1} (X(t) - tX(1))$$

and therefore of  $\lim_n P(D_n^+ < x)$ . Let  $Y(t) = X(t) - tX(1)$ . Then  $Y(t)$  is a Gaussian process with covariance

$$EY(t)Y(s) = s(1-t), \quad 0 \leq s \leq t \leq 1.$$

Consider the process

$$X^{(1)}(t) = (1+t)Y\left(\frac{t}{1+t}\right), \quad t \geq 0.$$

Its covariance is  $\min(s, t)$ , so  $X^{(1)}(t)$  is normalized Brownian motion. Therefore

$$\begin{aligned} P\left(\sup_{0 \leq t \leq 1} Y(t) \geq y\right) &= P\left(\sup_{t \geq 0} \frac{X^{(1)}(t)}{1+t} \geq y\right) \\ &= P\left(\sup_{t \geq 0} (X^{(1)}(t) - yt - y) \geq 0\right) \\ &= p(y, y) = e^{-2y^2}. \end{aligned}$$

The limiting distribution for  $D_n$  is similarly related to the probability of exiting from the two-sided region  $\{|x| < y(1+t)\}$ .

## Problems

### 5. Assuming

$$E \exp [-\lambda t_x^*] = \exp [-c\sqrt{\lambda} |x|],$$

find  $Ee^{-\lambda t^*}$ , where  $t^* = t^*(-1, +1)$ . Differentiating this with respect to  $\lambda$ , at  $\lambda = 0$ , find an expression for  $Et^*$  and compare this with the known value of  $Et^*$  to show that  $c = \sqrt{2}$ .

6. Use Wald's identity (see Problem 1) to get (13.23(1) and (2)) by using the equations for  $\lambda$  and  $-\lambda$ .

7. Using  $E \exp [-\lambda t_x^*] = \exp [-\sqrt{2\lambda} |x|]$ , prove that for  $x \geq 0$ ,

$$P\left(\sup_{0 \leq \tau \leq t} X(\tau) \geq x\right) = 2P(X(t) \geq x).$$

8. For  $S_1, S_2, \dots$  sums of independent, identically distributed random variables with zero means and finite second moments, find normalizing constants so that the following random variables converge in distribution to a nondegenerate limit, and evaluate the distribution of the limit, or the Laplace transform of the limit distribution

$$a) \max(S_1, \dots, S_n), \quad b) \max(|S_1|, \dots, |S_n|), \quad c) S_1 + \dots + S_n.$$

## 8. THE LAW OF THE ITERATED LOGARITHM

Let  $S_n, n = 1, 2, \dots$ , be successive sums of independent, identically distributed random variables  $Y_1, Y_2, \dots$ , with

$$EY_1 = 0, \quad EY_1^2 = \sigma^2 < \infty.$$

One version of the law of the iterated logarithm is

### Theorem 13.25

$$\overline{\lim} \frac{S_n}{\sqrt{2\sigma^2 n \log(\log n)}} = 1 \quad \text{a.s.}$$

Strassen [134] noted recently that even though this is a strong limit theorem, it follows from an invariance principle, and therefore is a distant consequence of the central limit theorem. The result follows fairly easily from the representation theorem, 13.8. What we need is

**Theorem 13.26.** *There is a probability space with a Brownian motion  $X(t)$  defined on it and a sequence  $\tilde{S}_n, n = 1, \dots$ , having the same distribution as  $S_n/\sigma, n = 1, \dots$ , such that*

$$(13.27) \quad \frac{\tilde{S}_{[t]} - X(t)}{\sqrt{2t \log(\log t)}} \rightarrow 0 \quad \text{a.s.}$$

as  $t \rightarrow \infty$ .

*Proof.* By 13.8, there is a sequence of independent, identically distributed, nonnegative random variables  $T_1, T_2, \dots, ET_1 = 1$  such that  $X(T_1 + \dots + T_n), n = 1, 2, \dots$ , has the same distribution as  $S_n/\sigma, n = 1, \dots$ . Therefore (13.27) reduces to proving that

$$\frac{X(T_1 + \dots + T_{[t]}) - X(t)}{\varphi(t)} \rightarrow 0 \quad \text{a.s.,}$$

where  $\varphi(t) = \sqrt{2t \log(\log t)}$ . By the law of large numbers,

$$\frac{T_1 + \dots + T_{[t]}}{t} \rightarrow 1 \quad \text{a.s.}$$



For any  $\epsilon > 0$ , there is an almost surely finite function  $t_0(\omega)$  such that for  $t \geq t_0(\omega)$ ,

$$T_1 + \cdots + T_{[t]} \in \left[ \frac{t}{1+\epsilon}, t(1+\epsilon) \right].$$

Let

$$M(t) = \sup_{\frac{t}{1+\epsilon} \leq \tau \leq (1+\epsilon)t} |X(\tau) - X(t)|.$$

Thus, for  $t_k = (1+\epsilon)^k$ , and  $t_k \leq t \leq t_{k+1}$ ,

$$\begin{aligned} M(t) &\leq \sup_{t_{k-1} \leq \tau \leq t_{k+2}} |X(\tau) - X(t)| \\ &\leq 2 \sup_{t_{k-1} \leq \tau \leq t_{k+2}} |X(\tau) - X(t_{k-1})|. \end{aligned}$$

In consequence, if we define

$$M_k = \sup_{t_{k-1} \leq \tau \leq t_{k+2}} |X(\tau) - X(t_{k-1})|,$$

$$\overline{\lim} \frac{M(t)}{\varphi(t)} \leq 2 \overline{\lim} \frac{M_k}{\varphi(t_k)}.$$

By 12.20,

$$PM_k(x) \leq 2P(|X(t_{k+2}) - X(t_{k-1})| > x).$$

Write  $t_{k+2} - t_{k-1} = \delta t_k$ , where  $\delta = (1+\epsilon)^2 - (1+\epsilon)^{-1}$ . Then

$$\begin{aligned} P(M_k \geq \sqrt{2\delta} \varphi(t_k)) &\leq 2P\left(\frac{|X(t_{k+2}) - X(t_{k-1})|}{\sqrt{t_{k+2} - t_{k-1}}} \geq 2\sqrt{\log(\log t_k)}\right) \\ &\leq \frac{2}{\sqrt{2\pi}} \exp[-2 \log(\log t_k)] = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{(\log t_k)^2} \\ &= \sqrt{\frac{2}{\pi}} \cdot \frac{1}{k^2 (\log(1+\epsilon))^2}. \end{aligned}$$

Use Borel-Cantelli again, getting

$$P(M_k \geq \sqrt{2\delta} \varphi(t_k) \text{ i.o.}) = 0$$

or

$$\overline{\lim} \frac{M_k}{\varphi(t_k)} \leq \sqrt{2\delta} \quad \text{a.s.}$$

Going back,

$$\overline{\lim} \frac{|S_{[t]} - X(t)|}{\varphi(t)} \leq \sqrt{8\delta} \quad \text{a.s.}$$

Taking  $\epsilon \downarrow 0$  gives  $\delta \rightarrow 0$  which completes the proof.

## 9. A MORE GENERAL INVARIANCE THEOREM

The direction in which generalization is needed is clear. Let the  $\{Y^{(n)}(t)\}$ ,  $t \in [0, 1]$ ,  $n = 0, 1, \dots$  be a sequence of processes such that  $Y^{(n)}(\cdot) \xrightarrow{\mathcal{D}} Y^{(0)}(\cdot)$  in the sense that all finite dimensional distributions converge to the appropriate limit. Suppose that all sample functions of  $\{Y^{(n)}(t)\}$  are in  $D$ . Suppose also that some metric  $\rho(\cdot, \cdot)$  is defined on  $D$ , and that in this metric  $H$  is a function on  $D$  a.s. continuous with respect to the distribution of  $Y^{(0)}(\cdot)$ . Find conditions to ensure that

$$H(Y^{(n)}(\cdot)) \xrightarrow{\mathcal{D}} H(Y^{(0)}(\cdot)).$$

This has been done for some useful metrics and we follow Skorokhod's strategy. The basic idea is similar to that in our previous work: Find processes  $\tilde{Y}^{(n)}(\cdot)$ ,  $n = 0, 1, \dots$  defined on a common probability space such that for each  $n$ ,  $\tilde{Y}^{(n)}(\cdot)$  has the same distribution as  $Y^{(n)}(\cdot)$ , and has all its sample functions in  $D$ . Suppose  $\tilde{Y}^{(n)}(\cdot)$  have the additional property that

$$\rho(\tilde{Y}^{(n)}(\cdot), \tilde{Y}^{(0)}(\cdot)) \xrightarrow{P} 0.$$

Then conclude, as in Section 5, that if  $H(Y^{(n)}(\cdot))$  and  $H(\tilde{Y}^{(n)}(\cdot))$  are random variables,

$$H(Y^{(n)}(\cdot)) \xrightarrow{\mathcal{D}} H(Y^{(0)}(\cdot)).$$

The basic tool is a construction that yields the very general

**Theorem 13.28.** *Let  $\{Y^{(n)}(t)\}$ ,  $t \in [0, 1]$ ,  $n = 0, 1, \dots$  be any sequence of processes such that  $Y^{(n)}(\cdot) \xrightarrow{\mathcal{D}} Y^{(0)}(\cdot)$ . Then for any countable set  $T \subset [0, 1]$ , there are processes  $\{\tilde{Y}^{(n)}(t)\}$ ,  $t \in T$ , defined on a common space such that*

- a) *For each  $n$ ,  $\{\tilde{Y}^{(n)}(t)\}$ ,  $t \in T$ , and  $\{Y^{(n)}(t)\}$ ,  $t \in T$ , have the same distribution.*
- b) *For every  $t \in T$ ,*

$$\tilde{Y}^{(n)}(t) \xrightarrow{\text{n.s.}} \tilde{Y}^{(0)}(t)$$

*as  $n \rightarrow \infty$ .*

*Proof.* The proof of this is based on some simple ideas but is filled with technical details. We give a very brief sketch and refer to Skorokhod [124] for a complete proof.

First, show that a single sequence of random variables  $X_n \xrightarrow{\mathcal{D}} X_0$  can be replaced by  $\tilde{X}_n \xrightarrow{\text{n.s.}} \tilde{X}_0$  on a common space with  $\tilde{X}_n \stackrel{\mathcal{D}}{=} X_n$ . It is a bit surprising to go from  $\xrightarrow{\mathcal{D}}$  to  $\xrightarrow{\text{n.s.}}$ . But if, for example,  $X_n$  are fair coin-tossing random variables such that  $X_n \xrightarrow{\mathcal{D}} X_0$ , then replace all  $X_n$  by the

random variables  $\tilde{X}_n$  on  $((0, 1), \mathcal{B}((0, 1)), dx)$  defined by

$$\tilde{X}_n(x) = \begin{cases} -1, & x \in (0, \frac{1}{2}], \\ +1, & x \in (\frac{1}{2}, 1). \end{cases}$$

Not only does  $\tilde{X}_n(x) \xrightarrow{\text{a.s.}} \tilde{X}_0(x)$ , but  $\tilde{X}_n(x) \equiv \tilde{X}_0(x)$ . In general, take  $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}) = ((0, 1), \mathcal{B}((0, 1)), dx)$ . The device is simple: If  $F_n(z)$ , the distribution function of  $X_n$ , is continuous with a unique inverse, then take

$$\tilde{X}_n(x) = F_n^{-1}(x).$$

Consequently,

$$\begin{aligned} \tilde{P}(\tilde{X}_n(x) < z) &= I\{x; F_n^{-1}(x) < z\} \\ &= I\{x; x < F_n(z)\} \\ &= F_n(z). \end{aligned}$$

Since  $X_n \xrightarrow{\mathcal{D}} X_0$ ,  $F_n(z) \rightarrow F_0(z)$  for every  $z$ ; thus  $F_n^{-1}(x) \rightarrow F_0^{-1}(x)$ , all  $x$ , or  $\tilde{X}_n(x) \xrightarrow{\text{a.s.}} \tilde{X}_0(x)$ . Because  $F_n$  may not have a unique inverse, define

$$\tilde{X}_n(x) = \inf_y \{y; F_n(y) > x\}.$$

Now verify that these variables do the job.

Generalize now to a sequence of process  $X_n = (X_1^{(n)}, \dots)$  such that  $X_n \xrightarrow{\mathcal{D}} X_0$ . Suppose we have a nice 1-1 mapping  $\theta: R^{(\infty)} \leftrightarrow B$ ,  $B \in \mathcal{B}_1$ , such that  $\theta, \theta^{-1}$  are measurable  $\mathcal{B}_\infty, \mathcal{B}_1(B)$  respectively, and such that the following holds:

$$\theta(X_n) \xrightarrow{\mathcal{D}} \theta(X_0).$$

Take  $Y_n, n \geq 0$ , random variables on a common space such that  $Y_n \stackrel{\mathcal{D}}{=} \theta(X_n)$  and  $Y_n \xrightarrow{\text{a.s.}} Y_0$ . Define  $\tilde{X}_n = \theta^{-1}(Y_n)$ . It is easy to see that  $X_n$  and  $\tilde{X}_n$  have the same distribution. If  $\theta^{-1}$  is smooth enough so that  $Y_n \xrightarrow{\text{a.s.}} Y_0$  implies that every coordinate of  $\theta^{-1}(Y_n)$  converges a.s. to the corresponding coordinate of  $\theta^{-1}(Y_0)$ , then this does it. To get such a  $\theta$ , let  $C_{n,k}$  be the set  $\{x; P(X_k^{(n)} = x) > 0\}$ . Let  $C = \bigcup_{n,k} C_{n,k}$ ;  $C$  is countable. Take  $\varphi(x): R^{(1)} \leftrightarrow (0, 1)$  to be 1-1 and continuous such that  $\varphi(C)$  contains no binary rationals. There is a 1-1 measurable mapping  $f: (0, 1)^{(\infty)} \leftrightarrow (0, 1)$  constructed in Appendix A.47. The mapping  $\theta: R^{(\infty)} \leftrightarrow (0, 1)$  defined by

$$\theta(x_1, x_2, \dots) = f(\varphi(x_1), \varphi(x_1), \varphi(x_2), \dots)$$

has all the necessary properties.

Take one more step. The process  $\{\tilde{X}^{(n)}(t)\}$ ,  $t \in T$ , having the same distribution as  $\{X^{(n)}(t)\}$ ,  $t \in T$ , has the property that a.s. every sample function is the restriction to  $T$  of a function in  $D$ . Take  $T$  dense in  $[0, 1]$ , and for any  $t \in [0, 1]$ ,  $t \notin T$ , define

$$\tilde{X}^{(n)}(t) = \lim_k \tilde{X}^{(n)}(t_k), \quad t_k \in T, t_k \downarrow t.$$

[Assume  $\{1\} \in T$ .] The processes  $\{\tilde{X}^{(n)}(t)\}$ ,  $t \in [0, 1]$ , defined this way have all their sample functions in  $D$ , except perhaps for a set of probability zero. Furthermore, they have the same distribution as  $X^{(n)}(\cdot)$  for each  $n$ .

Throwing out a set of probability zero, we get the statement: For each  $\omega$  fixed, the sample paths  $x_n(t) = \tilde{X}^{(n)}(t, \omega)$  are in  $D$  with the property that  $x_n(t) \rightarrow x_0(t)$  for all  $t \in T$ . The extra condition needed is something to guarantee that this convergence on  $T$  implies that  $\rho(x_n(\cdot), x_0(\cdot)) \rightarrow 0$  in the metric we are using. To illustrate this, use the metric

$$\rho(x(\cdot), y(\cdot)) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|,$$

introduced in the previous sections. Other metrics will be found in Skorokhod's article referred to above. Define

$$\delta_n(h) = \sup_{|t-\tau| \leq h} |x_n(t) - x_n(\tau)|, \quad \delta(h) = \overline{\lim}_n \delta_n(h).$$

If  $x_n(t) \rightarrow x_0(t)$  for  $t \in T$ , and  $x_0(t) \in D$ , then  $\lim_{h \downarrow 0} \delta(h) = 0$  implies  $\rho(x_n, x_0) \rightarrow 0$ . Hence the following:

**Theorem 13.29.** *Under the above assumptions, if*

$$(13.30) \quad \lim_{h \downarrow 0} \overline{\lim}_n P \left( \sup_{|t-\tau| \leq h} |Y^{(n)}(t) - Y^{(n)}(\tau)| \geq \epsilon \right) = 0,$$

then

$$\sup_{0 \leq t \leq 1} |\tilde{Y}^{(n)}(t) - \tilde{Y}^{(0)}(t)| \xrightarrow{P} 0.$$

*Proof.*  $\tilde{Y}^{(0)}(t)$  has continuous sample paths a.s. Because take  $T_m = \{t_1, \dots, t_m\} \subset T$ . Then, letting

$$M_n = \sup_{|t_j - t_k| \leq h} |\tilde{Y}^{(n)}(t_j) - \tilde{Y}^{(n)}(t_k)|,$$

we find that  $M_n \xrightarrow{\text{a.s.}} M_0$  follows. For  $\epsilon$  a continuity point in the distribution of  $M_0$ ,  $P(M_n \geq \epsilon) \rightarrow P(M_0 \geq \epsilon)$ . Therefore,

$$P(M_0 \geq \epsilon) \leq \overline{\lim}_n P \left( \sup_{|t-\tau| \leq h} |\tilde{Y}^{(n)}(t) - \tilde{Y}^{(n)}(\tau)| \geq \epsilon \right).$$

Letting  $T_m \uparrow T$  and using (13.30) implies the continuity.

Define random variables

$$\delta_n(h) = \sup_{|t-\tau| \leq h} |\tilde{Y}^{(n)}(t) - \tilde{Y}^{(n)}(\tau)|,$$

$$U_n = \sup_{0 \leq t \leq 1} |\tilde{Y}^{(n)}(t) - \tilde{Y}^{(0)}(t)|.$$

If  $T_m = \{t_1, \dots, t_m\} \subset T$  is such that

$$|t_{k+1} - t_k| \leq h, \quad k = 1, \dots, m-1,$$

then

$$U_n \leq \sup_{t \in T_m} |\tilde{Y}^{(n)}(t) - \tilde{Y}^{(0)}(t)| + \delta_n(h) + \delta_0(h).$$

The first term goes to zero a.s., leaving

$$\overline{\lim}_n P(U_n > \epsilon) \leq \overline{\lim}_n P(\delta_n(h) > \epsilon/3) + P(\delta_0(h) > \epsilon/3).$$

Take  $h \downarrow 0$ . Since  $U_n$  does not depend on  $h$ , the continuity of  $\tilde{Y}^{(0)}(\cdot)$ , and (13.30) yields

$$\overline{\lim}_n P(U_n > \epsilon) = 0 \Rightarrow U_n \xrightarrow{P} 0.$$

**Remark.** Note that under (13.30), since  $\tilde{Y}^{(0)}(\cdot)$  has continuous sample paths a.s., then  $Y^{(0)}(t)$  has a version with all sample paths continuous.

The general theorems are similar;  $Y^{(n)}(\cdot) \xrightarrow{\mathcal{D}} Y^{(0)}(\cdot)$  plus some equicontinuity condition on the  $Y^{(n)}(\cdot)$  gives

$$\rho(\tilde{Y}^{(n)}(\cdot), \tilde{Y}^{(0)}(\cdot)) \xrightarrow{P} 0.$$

**Problem 9.** For random variables having a uniform distribution on  $[0, 1]$ , and  $\hat{F}_n(x)$ , the sample distribution function defined in Section 6, use the multi-dimensional central limit theorem to show that

$$\{\hat{F}_n(\xi) - \xi\} \xrightarrow{\mathcal{D}} \{Y(\xi)\}, \quad \xi \in [0, 1],$$

where  $Y(\xi)$  is a Gaussian process with covariance

$$EY(\eta)Y(\xi) = \eta(1 - \xi), \quad 0 \leq \eta \leq \xi \leq 1.$$

Prove that (13.30) is satisfied by using 13.15 and the Skorokhod lemma 3.21.

## NOTES

The invariance principle as applied to sums of independent, identically distributed random variables first appeared in the work of Erdős and Kac [47, 1946] and [48, 1947]. The more general result of 13.12 is due to Donsker [30, 1951]. The method of imitating the sums by using a Brownian motion

evaluated at random times was developed by Skorokhod [126, 1961]. The possibility of using these methods on the Kolmogorov-Smirnov statistics was suggested by Doob [37] in a paper where he also evaluates the distribution of the limiting functionals on the Brownian motion. Donsker later [31, 1952] proved that Doob's suggested approach could be made rigorous. For some interesting material on the distribution of various functionals on Brownian motion, see Cameron and Martin [13], Kac [82], and Dinges [24].

Strassen's recent work [134, 1964] on the law of the iterated logarithm contains some fascinating generalizations of this law concerning the limiting fluctuations of Brownian motion. A relatively simple proof of the law of the iterated logarithm for coin-tossing is given by Feller [59, Vol. I]. A generalized version proved by Erdős [46, 1942] for coin-tossing, and extended by Feller [53, 1943] is: Let  $\varphi(n)$  be a positive, monotonically increasing function,  $S_n = Y_1 + \cdots + Y_n$ ,  $Y_1, \dots$ , independent and identically distributed random variables with mean zero and finite second moment. Then

$$P\left(\frac{S_n}{\sigma\sqrt{n}} > \varphi(n) \text{ i.o.}\right)$$

equals zero or one, depending on whether

$$\sum_1^{\infty} \frac{\varphi(n)}{n} e^{-(1/2)\varphi^2(n)}$$

converges or diverges.

The general question concerning convergence of a sequence of processes  $X^{(n)}(\cdot) \xrightarrow{D} X(\cdot)$  and related invariance results was dealt with in 1956 by Prokhorov [118] and by Skorokhod [124]. We followed the latter in Section 9.

The arc sine law has had an honorable history. Its importance in probability has been not so much in the theorem itself, as in the variety and power of the methods developed to prove it. For Brownian motion, it was derived by Paul Lévy [104, 1939]. Then Erdős and Kac [48, 1947] used an invariance argument to get it for sums of independent random variables. Then Sparre Andersen in 1954 [128] discovered a combinatorial proof that revealed the surprising fact that the law held for random variables whose second moments were not necessarily finite. Spitzer extended the combinatorial methods into entirely new areas [129, 1956]. For the latter, see particularly Spitzer's book [130], also the development by Feller [59, Vol. II]. Another interesting proof was given by Kac [82] for Brownian motion as a special case of a method that reduces the finding of distribution of functionals to related differential equations. There are at least three more proofs we know of that come from other areas of probability.

## CHAPTER 14

# MARTINGALES AND PROCESSES WITH STATIONARY, INDEPENDENT INCREMENTS

### 1. INTRODUCTION

In Chapter 12, Brownian motion was defined as follows:

- 1)  $X(t + \tau) - X(t)$  is independent of everything up to time  $t$ ,
- 2) The distribution of  $X(t + \tau) - X(t)$  depends only on  $\tau$ ,

$$3) \lim_{\Delta t \downarrow 0} \frac{P(|X(\Delta t)| \geq \epsilon)}{\Delta t} = 0.$$

The third assumption involved continuity and had the eventual consequence that a version of Brownian motion was available with all sample paths continuous.

If the third assumption is dropped, then we get a class of processes satisfying (1) and (2) which have the same relation to Brownian motion as the infinitely divisible laws do to the normal law. In fact, examining these processes gives much more meaning to the representation for characteristic functions of infinitely divisible laws.

These processes cannot have versions with continuous sample paths, otherwise the argument given in Chapter 12 forces them to be Brownian motion. Therefore, the extension problem that plagued us there and that we solved by taking a continuous version, comes back again. We deal with this problem in the same way—we take the smoothest possible version available. Of the results available relating to smoothness of sample paths, one of the most general is for continuous parameter martingale processes. So first we develop the martingale theorems. With this theory in hand, we then prove that there are versions of any of the processes satisfying (1) and (2) above, such that all sample paths are continuous except for jumps. Then we investigate the size and number of jumps in terms of the distribution of the process, and give some applications.

### 2. THE EXTENSION TO SMOOTH VERSIONS

Virtually all the well-known stochastic processes  $\{X(t)\}$ ,  $t \in I$ , can be shown to have versions such that all sample paths have only jump discontinuities. That is, the sample paths are functions  $x(t)$  which have finite right- and left-hand limits  $x(t-)$  and  $x(t+)$  at all  $t \in I$  for which these limits

can be defined. This last phrase refers to endpoints. Make the convention that if  $t$  is in the interior of  $I$ , both  $x(t-)$  and  $x(t+)$  limits can be defined. At a closed right endpoint, only the  $x(t-)$  limit can be defined. At a closed left endpoint only the  $x(t+)$  limit can be defined. At open (including infinite) endpoints, neither limit can be defined. We specialize a bit more and define:

**Definition 14.1.**  $D(I)$  is the class of all functions  $x(t)$ ,  $t \in I$ , which have only jump discontinuities and which are right-continuous; that is,  $x(t+) = x(t)$  for all  $t \in I$  such that  $x(t+)$  is defined.

Along with this goes

**Definition 14.2.** A process  $\{X(t)\}$ ,  $t \in I$  will be called continuous in probability from the right if whenever  $\tau \downarrow t$ ,

$$X(\tau) \xrightarrow{P} X(t).$$

We want to find conditions on the process  $\{X(t)\}$ ,  $t \in I$  so that a version exists with all sample paths in  $D(I)$ . As with Brownian motion, start by considering the variables of the process on a set  $T$  countable and dense in  $I$ , with the convention that  $T$  includes any closed endpoints of  $I$ . In the case of continuous sample paths the essential property was that for  $I$  finite, any function defined and uniformly continuous on  $T$  had an extension to a continuous function on  $I$ . The analog we need here is

**Definition 14.3.** A function  $x(t)$  defined on  $T$  is said to have only jump discontinuities in  $I$  if the limits

$$\lim_{s \downarrow t} x(s), \quad s \in T, \quad \lim_{s \uparrow t} x(s), \quad s \in T$$

exist and are finite for all  $t \in I$  where these limits can be defined.

**Proposition 14.4.** If  $x(t)$  defined on  $T$  has only jump discontinuities on  $I$ , then the function  $\tilde{x}(t)$  defined on  $I$  by

$$\tilde{x}(t) = \lim_{s \downarrow t} x(s), \quad s \in T$$

and  $\tilde{x}(b) = x(b)$  for  $b$  a closed right endpoint of  $I$  is in  $D(I)$ .

*Proof.* Let  $t_n \downarrow t$ ,  $t_n, t \in I$ , and take  $s_n \in T$ ,  $s_n > t_n > t$  such that  $s_n \downarrow t$  and  $x(s_n) - \tilde{x}(t_n) \rightarrow 0$ . Since  $x(s_n) \rightarrow \tilde{x}(t)$ , this implies  $\tilde{x}(t+) = \tilde{x}(t)$ . Now take  $t_n \uparrow t$ , and  $s_n \in T$  with  $t_n < s_n < t$  and  $\tilde{x}(t_n) - x(s_n) \rightarrow 0$ . This shows that  $\tilde{x}(t_n) \rightarrow \lim_{s \uparrow t} x(s)$ ,  $s \in T$ .

We use this to get conditions for the desired version.

**Theorem 14.5.** Let the process  $\{X(t)\}$ ,  $t \in I$ , be continuous in probability from the right. Suppose that almost every sample function of the countable process  $\{X(t)\}$ ,  $t \in T$ , has only jump discontinuities on  $I$ . Then there is a version of  $\{X(t)\}$ ,  $t \in I$ , with all sample paths in  $D(I)$ .



*Proof.* If for fixed  $\omega$ ,  $\{X(t, \omega)\}$ ,  $t \in T$ , does not have only jump discontinuities on  $I$ , put  $\tilde{X}(t, \omega) = 0$ , all  $t \in I$ . Otherwise, define

$$\tilde{X}(t, \omega) = \lim_{s \downarrow t} X(s, \omega), \quad s \in T$$

and  $\tilde{X}(b, \omega) = X(b, \omega)$  for  $b$  a closed right endpoint of  $I$ . By 14.4, the process  $\{\tilde{X}(t)\}$ ,  $t \in I$ , so defined has all its sample paths in  $D(I)$ . For any  $t \in I$  such that  $s_n \downarrow t$ ,  $s_n \in T$ ,  $\tilde{X}(t) = \lim_n X(s_n)$  a.s. By the continuity in probability from the right,  $X(s_n) \xrightarrow{P} X(t)$ . Hence

$$\tilde{X}(t) = X(t) \quad \text{a.s.,}$$

completing the proof.

**Problem 1.** For  $x(t) \in D(I)$ ,  $J$  any finite closed subinterval of  $I$ , show that

$$1) \sup_{t \in J} |x(t)| < \infty,$$

2) for any  $\delta > 0$ , the set

$$\{t \in J; |x(t) - x(t-)| \geq \delta\}$$

is finite.

3) The set of discontinuity points of  $x(t)$  is at most countable.

### 3. CONTINUOUS PARAMETER MARTINGALES

**Definition 14.6.** A process  $\{X(t)\}$ ,  $t \in I$ , is called a martingale (MG) if  $E|X(t)| < \infty$ , for all  $t \in I$ , and if for all  $s, t \in I$ ,  $s \leq t$ ,

$$E(X(t) | X(\tau), \tau \leq s) = X(s) \quad \text{a.s.}$$

Call the process a submartingale (SMG) if under the same conditions

$$E(X(t) | X(\tau), \tau \leq s) \geq X(s) \quad \text{a.s.}$$

This definition is clearly the immediate generalization of the discrete parameter case. The basic sample path property is:

**Theorem 14.7.** Let  $\{X(t)\}$ ,  $t \in I$  be a SMG. Then for  $T$  dense and countable in  $I$ , almost every sample function of  $\{X(t)\}$ ,  $t \in T$ , has only jump discontinuities on  $I$ .

*Proof.* It is sufficient to prove this for  $I$  a finite, closed interval  $[r, \tau]$ . Define

$$(14.8) \quad \begin{aligned} X^-(t-) &= \lim_{s \uparrow t} X(s), & X^+(t-) &= \overline{\lim_{s \uparrow t}} X(s), & s \in T, \\ X^-(t+) &= \lim_{s \downarrow t} X(s), & X^+(t+) &= \overline{\lim_{s \downarrow t}} X(s), & s \in T. \end{aligned}$$

Of course, the limits for  $r-$  and  $\tau+$  are not defined. First we show that for almost every sample function the limits in (14.8) are finite for all  $t \in I$ . In fact,

$$(14.9) \quad \sup_{t \in T} |X(t)| < \infty \quad \text{a.s.}$$

To show this, take  $T_N$  finite subsets of  $T$ ,  $T_N \uparrow T$  and  $r, \tau \in T_N$ . By adding together both (1) and (2) of 5.13, deduce that

$$(14.10) \quad P\left(\sup_{t \in T_N} |X(t)| > x\right) \leq \frac{1}{x} (2E|X(\tau)| - EX(r)).$$

Letting  $N \rightarrow \infty$  proves that (14.10) holds with  $T$  in place of  $T_N$ . Now take  $x \rightarrow \infty$  to prove (14.9).

Now assume all limits in (14.8) are finite. If a sample path of  $\{X(t)\}$ ,  $t \in T$  does not have only jump discontinuities on  $I$ , then there is a point  $v \in I$  such that either  $X^-(v-) < X^+(v-)$  or  $X^-(v+) < X^+(v+)$ . For any two numbers  $a < b$ , let  $D(a, b)$  be the set of all  $\omega$  such that there exists a  $v \in I$  with either

$$X^-(v-) < a < b < X^+(v-) \quad \text{or} \quad X^-(v+) < a < b < X^+(v+).$$

The union  $\cup D(a, b)$  over all rational  $a, b$ ;  $a < b$ , is then the set of all sample paths not having only jump discontinuities.

Take  $T_N$  finite subsets of  $T$  as above. Let  $\beta_N$  be the up-crossings of the interval  $[a, b]$  by the SMG sequence  $\{X(t_j)\}$ ,  $t_j \in T_N$  (see Section 4, Chapter 5). Then  $\beta_N \uparrow \beta$ , where  $\beta$  is a random variable, possibly extended. The significant fact is

$$D(a, b) \subset \{\beta = \infty\}.$$

Apply Lemma 5.17 to get

$$E\beta_N \leq \frac{E(X(\tau) - a)^+}{b - a}$$

to conclude that  $E\beta < \infty$ , hence  $P(D(a, b)) = 0$ . Q.E.D.

The various theorems concerning transformation of martingales by optional sampling and stopping generalize, if appropriate restrictions are imposed. See Doob [39, Chap. 7] for proofs under weak restrictions. We assume here that all the processes we work with have a version with all sample paths in  $D(I)$ .

**Proposition 14.11.** *Let  $\tau^*$  be a stopping time for a process  $\{X(t)\}$ ,  $t \in I$ , having all sample paths right-continuous. Then  $X(\tau^*)$  is a random variable.*

*Proof.* Approximate  $\tau^*$  by

$$\tau_n^* = \begin{cases} k/n, & (k-1)/n < \tau^* \leq k/n, & k > 1 \\ 1/n, & 0 \leq \tau^* \leq 1/n, & k = 1. \end{cases}$$

Then,  $X(t_n^*)$  is a random variable. For  $n$  running through  $2^m$ ,  $t_n^* \downarrow t^*$ , so by right-continuity,  $X(t_n^*) \rightarrow X(t^*)$  for every  $\omega$ .

We prove, as an example, the generalization of 5.31.

**Theorem 14.12.** *Let  $t^*$  be a stopping time for the SMG(MG)  $\{X(t)\}$ ,  $t \in [0, \infty)$ . If all the paths of the process are in  $D([0, \infty))$  and if*

$$\text{a) } E|X(t^*)| < \infty, \quad \text{b) } \lim_t \int_{\{t^* > t\}} |X(t)| dP = 0,$$

then

$$EX(t^*) \geq EX(0). \\ (=)$$

*Proof.* Suppose first that  $t^*$  is uniformly bounded,  $t^* \leq \tau$ . Take  $t_n^* \downarrow t^*$ ,  $t_n^* \leq \tau$ , but  $t_n^*$  taking on only a finite number of values. By 5.31,

$$EX(t_n^*) \geq EX(0). \\ (=)$$

and right-continuity implies  $X(t_n^*) \rightarrow X(t^*)$  everywhere. Some sort of boundedness condition is needed now to conclude  $EX(t_n^*) \rightarrow EX(t^*)$ . Uniform integrability is sufficient, that is,

$$\sup_n \int_{\{|X(t_n^*)| > x\}} |X(t_n^*)| dP$$

goes to zero as  $x \rightarrow \infty$ . If  $\{X(t)\}$  is a MG process, then  $\{|X(t)|\}$  is a SMG. Hence by the optional sampling theorem, 5.10,

$$(14.13) \quad \int_{\{|X(t_n^*)| > x\}} |X(t_n^*)| dP \leq \int_{\{|X(t_n^*)| > x\}} |X(\tau)| dP.$$

Let

$$M = \sup_{0 \leq t \leq \tau} |X(t)|.$$

By the right-continuity and (14.9),  $M < \infty$  a.s. Then the uniform integrability follows from  $E|X(\tau)| < \infty$  and

$$\int_{\{|X(t_n^*)| > x\}} |X(\tau)| dP \leq \int_{\{M > x\}} |X(\tau)| dP.$$

But if  $\{X(t)\}$  is not a MG, then use this argument: If the SMG  $\{X(t)\}$   $t \geq 0$ , were bounded below, say,  $X(t) \geq \alpha$ , all  $t \leq \tau$  and  $\omega$ , then for  $x > |\alpha|$ ,

$$\begin{aligned} \int_{\{|X(t_n^*)| > x\}} |X(t_n^*)| dP &= \int_{\{|X(t_n^*)| > x\}} X(t_n^*) dP \\ &\leq \int_{\{|X(t_n^*)| > x\}} X(\tau) dP \leq \int_{\{|X(t_n^*)| > x\}} |X(\tau)| dP. \end{aligned}$$

This gets us to (14.13) again. Proceed as above to conclude  $EX(t^*) \geq EX(0)$ . In general, for  $\alpha$  negative, take  $Y(t) = \max(\alpha, X(t))$ . Then  $\{Y(t)\}$ ,  $t \geq 0$  is a SMG bounded below, so  $EY(t^*) \geq EY(0)$ . Take  $\alpha \rightarrow -\infty$  and note that

$$EX(t^*) - EY(t^*) = \int_{\{X(t^*) \leq \alpha\}} (X(t^*) - \alpha) dP$$

goes to zero. Similarly for  $EX(0) - EY(0)$ , proving the theorem for bounded stopping times. If  $t^*$  is not bounded, define the stopping time  $t^{**}$  as  $\min(t^*, \tau)$ . Then

$$E|X(t^{**})| \leq E|X(t^*)| + E|X(\tau)| < \infty,$$

so  $EX(t^{**}) \geq EX(0)$ . But

$$EX(t^*) - EX(t^{**}) = \int_{\{t^* > \tau\}} (X(t^*) - X(\tau)) dP.$$

The first term in this integral goes to zero as  $\tau \rightarrow \infty$  because  $E|X(t^*)| < \infty$  by hypothesis. For the second term, simply take a sequence  $\tau_n \rightarrow \infty$  such that

$$\int_{\{t^* > \tau_n\}} |X(\tau_n)| dP \rightarrow 0.$$

For this sequence  $EX(t^{**}) \rightarrow EX(t^*)$ , completing the proof.

**Problem 2.** For Brownian motion  $\{X(t)\}$ ,  $t \geq 0$ , and  $t^* = t^*(a, b)$ , prove using 14.12 that  $EX(t^*) = 0$ ,  $EX^2(t^*) = Et^*$ .

#### 4. PROCESSES WITH STATIONARY, INDEPENDENT INCREMENTS

**Definition 14.14.** A process  $\{X(t)\}$ ,  $t \in [0, \infty)$ , has independent increments if for any  $t$  and  $\tau > 0$ ,  $\mathcal{F}(X(t + \tau) - X(t))$  is independent of  $\mathcal{F}(X(s), s \leq t)$ .

The stationary condition is that the distribution of the increase does not depend on the time origin.

**Definition 14.15.** A process  $\{X(t)\}$ ,  $t \in [0, \infty)$  is said to have stationary increments if  $\mathcal{L}(X(t + \tau) - X(t))$ ,  $\tau \geq 0$ , does not depend on  $t$ .

In this section we will deal with processes having independent, stationary increments, and we further normalize by taking  $X(0) \equiv 0$ . Note that

$$\begin{aligned} X(t) &= (X(t) - X(t - \tau)) \\ &\quad + (X(t - \tau) - X(t - 2\tau)) + \cdots + (X(\tau) - X(0)), \end{aligned}$$

where  $\tau = t/n$ . Or

$$X(t) = X_1^{(n)} + \cdots + X_n^{(n)},$$

where the  $X_k^{(n)}$ ,  $k = 1, \dots, n$  are independent and identically distributed. Ergo,  $X(t)$  must have an infinitely divisible distribution. Putting this formally, we have the following proposition.

**Proposition 14.16.** *Let  $\{X(t)\}$ ,  $t \in [0, \infty)$  be a process with independent, stationary increments; then  $X(t)$  has an infinitely divisible distribution for every  $t \geq 0$ .*

It follows from  $X(t+s) = (X(t+s) - X(s)) + (X(s) - X(0))$  that  $f_t(u)$ , the characteristic function of  $X(t)$ , satisfies the identity

$$(14.17) \quad f_{t+s}(u) = f_t(u)f_s(u).$$

If  $f_t(u)$  had any reasonable smoothness properties, such as  $f_t(u)$  measurable  $\mathfrak{B}_1$  in  $t$  for each  $u$ , then (14.17) would imply that  $f_t(u) = [f_1(u)]^t$ ,  $t \geq 0$ . Unfortunately, a pathology can occur: Let  $\varphi(t)$  be a real solution of the equation  $\varphi(t+s) = \varphi(t) + \varphi(s)$ ,  $t, s \geq 0$ . Nonlinear solutions of this do exist [33]. They are nonmeasurable and unbounded in every interval. Consider the degenerate process  $X(t) \equiv \varphi(t)$ . This process has stationary, independent increments.

Starting with any process  $\{X(t)\}$  such that  $f_t(u) = [f_1(u)]^t$ , then the process  $X'(t) = X(t) + \varphi(t)$  has also stationary, independent increments, but  $f'_t(u) \neq [f'_1(u)]^t$ . This is the extent of the pathology, because it follows from Doob [39, pp. 407 ff.] that if  $\{X(t)\}$  is a process with stationary, independent increments, then there is a function  $\varphi(t)$ ,  $\varphi(t+s) = \varphi(t) + \varphi(s)$  such that the process  $\{X^{(1)}(t)\}$ ,  $X^{(1)}(t) = X(t) - \varphi(t)$  has stationary, independent increments,  $X^{(1)}(0) \equiv 0$ , and  $f^{(1)}_t(u)$  is continuous in  $t$  for every  $u$ . Actually, this is not difficult to show directly in this case (see Problem 3). A sufficient condition that eliminates this unpleasant case is given by

**Proposition 14.18.** *Let  $\{X(t)\}$  be a process with stationary, independent increments such that  $f_t(u)$  is continuous at  $t = 0$  for every  $u$ ; then  $\{X(t)\}$  is continuous in probability and  $f_t(u) = [f_1(u)]^t$ .*

*Proof.* Fix  $u$ , then taking  $s \downarrow 0$  in the equations  $f_{t+s}(u) = f_t(u)f_s(u)$  and  $f_t(u) = f_{t-s}(u)f_s(u)$  proves  $f_t(u)$  continuous for all  $t \geq 0$ . The only continuous solutions of this functional equation are the exponentials. Therefore  $f_t(u) = e^{t\psi(u)}$ . Evaluate this at  $t = 1$  to get  $f_1(u) = e^{\psi(u)}$ . Use  $\lim_{s \downarrow 0} f_s(u) = 1$  to conclude  $X(s) \xrightarrow{\mathcal{D}} 0$  as  $s \rightarrow 0$ , implying  $X(s) \xrightarrow{P} 0$ . Since

$$\mathbb{E}(X(t) - X(t-s)) = \mathbb{E}(X(t+s) - X(t)) = \mathbb{E}(X(s)), \quad s \geq 0,$$

the proposition follows.

The converse holds.

**Proposition 14.19.** *Given the characteristic function  $f(u)$  of an infinitely divisible distribution, there is a unique process  $\{X(t)\}$  with stationary, independent increments,  $X(0) \equiv 0$ , such that  $f_t(u) = [f(u)]^t$ .*

**Remark.** In the above statement, uniqueness means that every process satisfying the given conditions has the same distribution.

*Proof.* All that is necessary to prove existence is to produce finite-dimensional consistent distribution functions. To specify the distribution of  $X(t_1), X(t_2), \dots, X(t_n)$ ,  $t_1 < t_2 < \dots < t_n$ , define variables  $Y_1, Y_2, \dots, Y_n$  as being independent,  $Y_k$  having characteristic function  $[f(u)]^{t_k - t_{k-1}}$ ,  $t_0 = 0$ . Define

$$\mathbb{E}(X(t_1), \dots, X(t_n)) = \mathbb{E}(Y_1, Y_1 + Y_2, \dots, Y_1 + Y_2 + \dots + Y_n).$$

The consistency is obvious. By the extension theorem for processes, 12.14, there is a process  $\{X(t)\}$ ,  $t \geq 0$ , having the specified distributions. For these distributions,  $X(t + \tau) - X(t)$  is independent of the vector  $(X(t_1), \dots, X(t_n))$ ,  $t_1, \dots, t_n \leq t$ . Thus the process  $\{X(t)\}$  has independent increments. Further, the characteristic function of  $X(t + \tau) - X(t)$  is  $[f(u)]^\tau$ , all  $t \geq 0$ , so implying stationarity of the increments. Of course, the characteristic function of  $X(t)$  is  $[f(u)]^t$ . By construction,  $X(0) = 0$  a.s. since  $f_0(u) = 1$ , so we take a version with  $X(0) \equiv 0$ . If there is any other such process  $\{X(t)\}$  having characteristic function  $[f(u)]^t$ , clearly its distribution is the same.

Since there is a one-to-one correspondence between processes with stationary, independent increments, continuous in probability and characteristic functions of infinitely divisible distributions, we add the terminology: If

$$Ee^{iuX(t)} = e^{t\psi(u)},$$

call  $\psi(u)$  the *exponent function* of the process.

### Problems

3. Since  $|f_t(u)| \leq 1$ , all  $u$ , show that (14.17) implies that  $|f_t(u)| = |f_1(u)|^t$ . By (9.20),  $\log f_t(u) = i\beta(t)u + \int \varphi(x, u)\gamma_t(dx)$ . Use  $|f_t(u)| = |f_1(u)|^t$  to show that  $\gamma_t(R^{(1)})$  is uniformly bounded in every finite  $t$ -interval. Deduce that  $\gamma_{t+s} \equiv \gamma_t + \gamma_s$ ; hence show that  $\gamma_t \equiv t\gamma_1$ . Now  $\beta(t + s) = \beta(t) + \beta(s)$  by the unique representation. Conclude that  $X^{(1)}(t) = X(t) - \beta(t)$  has a continuous characteristic function  $f_t^{(1)}(u)$ . (It is known that every measurable solution of  $\varphi(t + s) = \varphi(t) + \varphi(s)$  is linear. Therefore if  $\beta(t)$  is measurable, then it is continuous and  $f_t(u)$  is therefore continuous.)

4. For a process with independent increments show that

$$\mathcal{F}(X(t + \tau) - X(t))$$

independent of  $\mathcal{F}(X(s), s \leq t)$  for each  $t, \tau \Rightarrow \mathcal{F}(X(t + \tau) - X(t), \tau \geq 0)$  independent of  $\mathcal{F}(X(s), s \leq t)$ .

5. For a process with stationary, independent increments, show that for  $B \in \mathcal{B}_1$

$$1) \quad P(X(t + \tau) \in B \mid X(s), s \leq t) = P(X(t + \tau) \in B \mid X(t)) \quad \text{a.s.},$$

$$2) \quad P(X(t + \tau) \in B \mid X(t) = x) = P(X(\tau) \in B - x) \quad \text{a.s.} \quad P(X(t) \in dx).$$

## 5. PATH PROPERTIES

We can apply the martingale results to get this theorem:

**Theorem 14.20.** *Let  $\{X(t)\}$ ,  $t \geq 0$  be a process with stationary, independent increments continuous in probability. Then there is a version of  $\{X(t)\}$  with all sample paths in  $D([0, \infty))$ .*

**Remark.** There are a number of ways to prove this theorem. It can be done directly, using Skorokhod's lemma (3.21) in much the same way as was done for path continuity in Brownian motion. But since we have martingale machinery available, we use a device suggested by Doob [39].

*Proof.* Take, as usual,  $X(0) \equiv 0$ . If  $E|X(t)| < \infty$ , all  $t \geq 0$ , we are finished. Because, subtracting off the means  $EX(t)$  if necessary, assume  $EX(t) \equiv 0$ . Then  $\{X(t)\}$ ,  $t \geq 0$  is a martingale since for  $0 \leq s \leq t$ ,

$$E(X(t) | X(\tau), \tau \leq s) = E(X(t) - X(s) | X(\tau), \tau \leq s) + X(s) = X(s) \quad \text{a.s.}$$

Simply apply the martingale path theorem (14.7), the continuity in probability, and 14.5.

If  $E|X(t)|$  is not finite for all  $t$ , one interesting proof is the following: Take  $\varphi(x)$  any continuous bounded function. The process on  $[0, \tau]$  defined by

$$Y(t) = E(\varphi(X(\tau)) | X(s), s \leq t)$$

is a martingale. But (see Problem 5),

$$Y(t) = E(\varphi(X(\tau)) | X(t)) = \theta(\tau - t, X(t)),$$

where

$$\theta(t, x) = E\varphi(X(t) + x).$$

The plan is to deduce the path properties of  $X(t)$  from the martingale path theorem by choosing suitable  $\varphi(x)$ . Take  $\varphi(x)$  strictly increasing,  $\varphi(+\infty) = \alpha$ ,  $\varphi(-\infty) = -\alpha$ ,  $E|\varphi(X(\tau))| = 1$ . By the continuity in probability of  $X(t)$ ,  $\theta(t, x)$  is continuous in  $t$ . It is continuous and strictly increasing in  $x$ , hence jointly continuous in  $x$  and  $t$ . Thus,

$$y = \theta(\tau - t, x)$$

has an inverse

$$x = \Theta(t, y)$$

jointly continuous in  $(t, y)$ , for  $0 \leq t \leq \tau$ ,  $|y| < \alpha$ , and

$$X(t) = \Theta(t, Y(t)).$$

For  $T$  dense and countable in  $[0, \tau]$ , the martingale path theorem implies that

$\{Y(t)\}$ ,  $t \in T$ , has only jump discontinuities on  $[0, \tau]$ . Hence, for all paths such that  $\sup_{t \in T} |Y(t)| < \alpha$ ,  $\{X(t)\}$ ,  $t \in T$ , has only jump discontinuities on  $[0, \tau]$ . Now, to complete the proof, we need to show that  $\sup_{t \in T} |X(t)| < \infty$ , a.s. because for every sample function,

$$\sup_{t \in T} |X(t)| = \infty \Leftrightarrow \sup_{t \in T} |Y(t)| = \alpha.$$

Since  $|Y(t)|$  is a SMG, apply 5.13 to conclude

$$P\left(\sup_{t \in T} |Y(t)| > \frac{\alpha}{2}\right) \leq \frac{2}{\alpha} E |Y(\tau)|.$$

Since

$$E |Y(\tau)| = E |\varphi(X(\tau))| = 1,$$

we get

$$P\left(\sup_{t \in T} |X(t)| = \infty\right) \leq \frac{2}{\alpha}.$$

Since  $\alpha$  can be made arbitrarily large, conclude that

$$P\left(\sup_{t \in T} |X(t)| = \infty\right) = 0. \quad \text{Q.E.D.}$$

In the rest of this chapter we take all processes with stationary, independent increments to be continuous in probability with sample paths in  $D([0, \infty))$ .

### Problems

6. Prove, using Skorokhod's lemma 3.21 directly, that

$$\sup_{t \in T} |X(t)| < \infty \quad \text{a.s.}$$

7. For  $\{X(t)\}$ ,  $t \geq 0$ , a process with stationary, independent increments with sample paths in  $D([0, \infty))$ , show that the strong Markov property holds; that is, if  $t^*$  is any stopping time, then

$$\{X(t + t^*) - X(t^*), \quad t \geq 0,$$

has the same distribution as  $\{X(t)\}$ ,  $t \geq 0$  and is independent of  $\mathcal{F}(X(t), t \leq t^*)$ .

8. For  $\{X(t)\}$ ,  $t \geq 0$ , a process continuous in probability with sample paths in  $D([0, \infty))$ , show that

$$P(|X(t) - X(t-)| > 0) = 0$$

for all  $t \geq 0$ .



## 6. THE POISSON PROCESS

This process stands at the other end of the spectrum from Brownian motion and can be considered as the simplest and most basic of the processes with stationary, independent increments. We get at it this way: A sample point  $\omega$  consists of any countable collection of points of  $[0, \infty)$  such that if  $N(I, \omega)$  is the number of points of  $\omega$  falling into the interval  $I$ , then  $N(I, \omega) < \infty$  for all finite intervals  $I$ . Define  $\mathcal{F}$  on  $\Omega$  such that all  $N(I, \omega)$  are random variables, and impose on the probability  $P$  these conditions.

### Conditions 14.21

- The number of points in nonoverlapping intervals is independent. That is,  $I_1, \dots, I_k$  disjoint  $\Rightarrow N(I_1), \dots, N(I_k)$  independent.*
- The distribution of the number of points in any interval  $I$  depends only on the length  $\|I\|$  of  $I$ .*

A creature with this type of sample space is called a point process. Conditions (a) and (b) arise naturally under wide circumstances: For example, consider a Geiger counter held in front of a fairly sizeable mass of radioactive material, and let the points of  $\omega$  be the successive registration times. Or consider a telephone exchange where we plot the times of incoming telephone calls over a period short enough so that the disparity between morning, afternoon, and nighttime business can be ignored. Define

$$(14.22) \quad X(t) = N((0, t]).$$

Then by 14.21(a) and (b),  $X(t)$  is a process with stationary, independent increments. Now the question is: *Which one?* Actually, the prior part of this question is: *Is there one?* The answer is:

**Theorem 14.23.** *A process  $X(t)$  with stationary, independent increments has a version with all sample paths constant except for upward jumps of length one if and only if there is a parameter  $\lambda \geq 0$  such that*

$$E e^{iuX(t)} = e^{\lambda t(e^{iu} - 1)}.$$

**Remark.** By expanding, we find that  $X(t)$  has the Poisson distribution

$$P(X(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

and

$$P(N(I) = n) = \frac{(\lambda \|I\|)^n}{n!} e^{-\lambda \|I\|}.$$

*Proof.* Let  $X(t)$  be a process with the given characteristic function, with paths in  $D([0, \infty))$ . Then

$$P(X(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

that is,  $X(t)$  is concentrated on the nonnegative integers  $I^+$ , so that  $P(X(t) \in I^+, \text{ all } t \in T) = 1$  for any countable set  $T$ . Taking  $T$  dense in  $[0, \infty)$  implies that the paths of the process are integer-valued, with probability one. Also,  $\{X(t)\}$ ,  $t \in T$ , has nondecreasing sample paths, because  $\mathbb{E}(X(t + \tau) - X(t)) = \mathbb{E}(X(\tau))$ , and  $X(\tau) \geq 0$ , a.s. Therefore there is a version of  $X(t)$  such that all sample paths take values in  $I^+$  and jump upward only. I want to show that for this version,

$$P\left(\sup_{0 \leq t \leq 1} (X(t) - X(t-)) \leq 1\right) = 1.$$

By Problem 8, the probability that  $X(t) - X(t-) > 0$  for  $t$  rational is zero. Hence, a.s.,

$$\sup_{0 \leq t \leq 1} (X(t) - X(t-)) = \lim_n \max_{1 \leq k \leq n} (X(k/n) - X(k - 1/n)).$$

But

$$\begin{aligned} P\left(\max_{1 \leq k \leq n} (X(k/n) - X(k - 1/n)) \leq 1\right) &= P(X(1/n) \leq 1)^n \\ &= [e^{-\lambda/n}(1 + \lambda/n)]^n \\ &\rightarrow 1 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

To go the other way, let  $X(t)$  be any process with stationary, independent increments, integer-valued, such that  $X(t) - X(t-) = 0$  or 1. Take  $t_1^*$  to be the time until the first jump. It is a stopping time. Take  $t_2^*$  to be the time until the first jump of  $X(t + t_1^*) - X(t_1^*)$ . We know that  $t_2^*$  is independent of  $t_1^*$  with the same distribution. The time until the  $n$ th jump is  $t_1^* + \cdots + t_n^*$ ,

$$P(t_1^* > t + \tau) = P(t_1^* > \tau, X(\xi) - X(\tau) = 0, \tau \leq \xi \leq t + \tau).$$

Now  $\{t_1^* > \tau\} = \{t_1^* \leq \tau\}^c$ , hence  $\{t_1^* > \tau\}$  is in  $\mathcal{F}(X(t), t \leq \tau)$ . Therefore

$$\begin{aligned} P(t_1^* > t + \tau) &= P(t_1^* > \tau)P(X(\xi) - X(\tau) = 0, \tau \leq \xi \leq t + \tau) \\ &= P(t_1^* > \tau)P(t_1^* > t). \end{aligned}$$

This is the exponential equation; the solution is

$$P(t_1^* > t) = e^{-\lambda t}$$

for some  $\lambda \geq 0$ . To finish, write

$$P(X(t) = n) = P(t_1^* + \cdots + t_n^* < t) - P(t_1^* + \cdots + t_{n+1}^* < t).$$

Denote  $Q_n(t) = P(t_1^* + \cdots + t_n^* > t)$ , and note that

$$P(t_1^* + \cdots + t_{n+1}^* > t \mid t_1^* + \cdots + t_n^* = \xi) = \begin{cases} e^{-\lambda(t-\xi)}, & t \geq \xi, \\ 1, & t < \xi. \end{cases}$$

So we have

$$Q_{n+1}(t) = Q_n(t) - e^{-\lambda t} \int_0^t e^{\lambda \xi} Q_n(d\xi), \quad n \geq 1.$$

This recurrence relation and  $Q_1(t) = e^{-\lambda t}$  gives

$$Q_{n+1}(t) = e^{-\lambda t} \left( 1 + \lambda t + \cdots + \frac{(\lambda t)^n}{n!} \right),$$

leading to

$$P(X(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \quad \text{Q.E.D.}$$

## 7. JUMP PROCESSES

We can use the Poisson processes as building blocks. Let the jump points of a Poisson process  $Y(t)$  with parameter  $\lambda > 0$  be  $t_1, t_2, \dots$ . Construct a new process  $X(t)$  by assigning jump  $X_1$  at time  $t_1$ ,  $X_2$  at time  $t_2, \dots$ , where  $X_1, X_2, \dots$  are independent, identically distributed random variables with distribution function  $F(x)$ , and  $\mathcal{F}(X)$  is independent of  $\mathcal{F}(Y(t), t \geq 0)$ . Then

**Proposition 14.24.**  $X(t)$  is a process with stationary, independent increments, and

$$E \exp[iuX(t)] = \exp \left\{ \lambda t \int [\exp(iux) - 1] dF(x) \right\}.$$

*Proof.* That  $X(t)$  has stationary, independent increments follows from the construction. Let  $h(u) = \int e^{iux} dF$ . Note that

$$E(\exp[iuX(t)] | Y(t) = n) = E \exp[iu(X_1 + \cdots + X_n)] = [h(u)]^n,$$

implying

$$E e^{iuX(t)} = \sum_0^\infty [h(u)]^n \frac{(\lambda t)^n e^{-\lambda t}}{n!} = e^{\lambda t(h(u)-1)}.$$

In Chapter 9, Section 4, it was pointed out that for a generalized Poisson distribution with jumps of size  $x_1, x_2, \dots$ , each jump size contributes an independent component to the distribution. This is much more graphically illustrated here. Say that a process has a jump of size  $B \in \mathcal{B}_1$  at time  $t$  if  $X(t) - X(t-) \in B$ . A process with stationary, independent increments and exponent function

$$\int (e^{iux} - 1) \mu(dx)$$

with  $\mu(R^{(1)}) < \infty$  is of the type treated above, with  $\lambda = \mu(R^{(1)})$ ,  $F(B) = \mu(B)/\mu(R^{(1)})$ . Therefore it has sample paths constant except for jumps. Let

$R^{(1)} - \{0\} = \bigcup_{k=1}^n B_k$ , for disjoint  $B_k \in \mathcal{B}_1$ ,  $k = 1, \dots, n$ . Define processes  $\{X(B_k, t)\}$ ,  $t \geq 0$ , by

$$X(B_k, t) = \sum_{\tau \leq t} \chi_{B_k}(X(\tau) - X(\tau -)).$$

Thus  $X(B_k, t)$  is the sum of the jumps of  $X(t)$  of size  $B_k$  up to time  $t$ . We need to show measurability of  $X(B_k, t)$ , but this follows easily if we construct the process  $X(t)$  by using a Poisson process  $Y(t)$  and independent jump variables  $X_1, X_2, \dots$ , as above. Then, clearly,

$$X(t) = \sum_{k=1}^n X(B_k, t).$$

Now we prove:

**Proposition 14.25.** *The processes  $\{X(B_k, t)\}$ ,  $t \geq 0$ , are independent of each other, and  $\{X(B_k, t)\}$ ,  $t \geq 0$  is a process with stationary, independent increments and exponent function*

$$\int_{B_k} (e^{iux} - 1) \mu(dx).$$

*Proof.* It is possible to prove this directly, but the details are messy. So we resort to an indirect method of proof. Construct a sample space and processes  $\{\tilde{X}^{(k)}(t)\}$ ,  $t \geq 0$ ,  $n = 1, \dots, n$  which are independent of each other and which have stationary, independent increments with exponent functions

$$\int_{B_k} (e^{iux} - 1) \mu(dx).$$

Each of these has the same type of distribution as the processes of 14.24, with  $\lambda = \mu(B_k)$ ,  $F_k(B) = \mu(B \cap B_k)/\mu(B_k)$ . Hence  $F_k(dx)$  is concentrated on  $B_k$  and  $\{\tilde{X}^{(k)}(t)\}$ ,  $t \geq 0$ , has jumps only of size  $B_k$ . Let

$$\tilde{X}(t) = \sum_{k=1}^n \tilde{X}^{(k)}(t).$$

Then  $\{\tilde{X}(t)\}$ ,  $t \geq 0$  has stationary, independent increments and the same characteristic function for every  $t$  as  $X(t)$ . Therefore  $\{\tilde{X}(t)\}$  and  $\{X(t)\}$  have the same distribution. But by construction,

$$\tilde{X}^{(k)}(t) = \sum_{\tau \leq t} \chi_{B_k}(\tilde{X}(t) - \tilde{X}(t -)),$$

therefore  $\{\tilde{X}^{(k)}(t)\}$ ,  $t \geq 0$  is defined on the  $\{\tilde{X}(t)\}$  process by the same function as  $\{X(B_k, t)\}$  on the  $\{X(t)\}$  process. Hence the processes  $\{X(B_k, t)\}$ ,  $\{X(t)\}$  have the same joint distribution as  $\{\tilde{X}^{(k)}(t)\}$ ,  $\{\tilde{X}(t)\}$ , proving the proposition.

## 8. LIMITS OF JUMP PROCESSES

The results of the last section give some insight into the description of the paths of the general process. First, let  $\{X(t)\}$ ,  $t \geq 0$ , be a process with stationary, independent increments, whose exponent function  $\psi$  has only an integral component:

$$\psi(u) = \int_{\{0\}^c} \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \mu(dx).$$

If  $\mu(R^{(1)}) < \infty$ , then the process is of the form  $\{X(t) - \beta t\}$ , where  $\{X(t)\}$ ,  $t \geq 0$  is of the type studied in the previous section, with sample paths constant except for isolated jumps. In the general case,  $\mu$  assigns infinite mass to arbitrarily small neighborhoods of the origin. This leads to the suspicion that the paths for these processes have an infinite number of jumps of very small size. To better understand this, let  $D$  be any neighborhood of the origin,  $\{X(D^c, t)\}$ ,  $t \geq 0$ , be the process of jumps of size greater than  $D$ , where we again define: *the process  $\{X(B, t)\}$ ,  $t \geq 0$ , of jumps of  $\{X(t)\}$  of size  $B \in \mathcal{B}_1$ ,  $\{0\} \notin B$ , is given by*

$$(14.26) \quad X(B, t) = \sum_{\tau \leq t} \chi_B(X(t) - X(t-)).$$

Assuming that the results analogous to the last section carry over,  $\{X(D^c, t)\}$  has exponent function

$$\int_{D^c} (e^{iux} - 1) \mu(dx).$$

Letting

$$\beta = \int_{D^c} \frac{x}{1+x^2} \mu(dx),$$

we have that the exponent function of  $X(D^c, t) - \beta t$  is given by

$$\psi_D(u) = \int_{D^c} \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \mu(dx).$$

Therefore, as  $D$  shrinks down to  $\{0\}$ ,  $\psi_D(u) \rightarrow \psi(u)$ , and in some sense we expect that  $X(t)$  is the limit of  $X(D^c, t) - \beta t$ . In fact, we can get a very strong convergence.

**Theorem 14.27.** *Let  $\{X(t)\}$ ,  $t \geq 0$ , be a process with stationary, independent increments having only an integral component in its characteristic function. Take  $\{D_n\}$  neighborhoods of the origin such that  $D_n \downarrow \{0\}$ , then  $\{X(D_n^c, t)\}$ ,  $t \geq 0$  is a process with stationary, independent increments and exponent function*

$$\int_{D_n^c} (e^{iux} - 1) \mu(dx).$$

Put

$$\beta_n = \int_{D_n^c} \frac{x}{1+x^2} \mu(x).$$

For any  $t_0 < \infty$ ,

$$\sup_{0 \leq t \leq t_0} |X(t) - X(D_n^c, t) + \beta_n t| \xrightarrow{P} 0$$

as  $n \rightarrow \infty$ .

*Proof.* Take  $D_1 = R^{(1)}$ ,  $B_n = D_{n+1} - D_n$ . Construct a probability space on which there are processes  $\{\tilde{Z}_n(t)\}$ ,  $t \geq 0$ , with stationary, independent increments such that the processes are independent of each other, with exponent functions

$$\psi_n(u) = \int_{B_n} (e^{iux} - 1) \mu(dx).$$

Then 14.25 implies that the paths of  $\tilde{Z}_n(t)$  are constant except for jumps of size  $B_n$ . Denote

$$\tilde{X}_n(t) = \sum_1^n \tilde{Z}^{(k)}(t) \quad \text{and} \quad b_n = \int_{B_n} \frac{x}{1+x^2} \mu(dx).$$

Consequently,  $\tilde{X}_n(t) - \beta_n t$  is the sum of the independent components  $\tilde{Z}_k(t) - b_k t$ . Since the characteristic functions converge, for every  $t$ ,

$$\tilde{X}_n(t) - \beta_n t \xrightarrow{\mathcal{D}}.$$

Because we are dealing with sums of independent random variables, 8.36 implies that for every  $t$  there is a random variable  $\tilde{X}(t)$  such that

$$\tilde{X}_n(t) - \beta_n t \xrightarrow{\text{a.s.}} \tilde{X}(t).$$

This implies that  $\{\tilde{X}(t)\}$ ,  $t \geq 0$ , is a process with stationary, independent increments having the same distribution as  $\{X(t)\}$ ,  $t \geq 0$ . We may assume that  $\{\tilde{X}(t)\}$ ,  $t \geq 0$  has all its sample paths in  $D([0, \infty))$ . Take, for example,  $t_0 = 1$ ,  $T$  any set dense in  $[0, 1]$ . For  $\tilde{Y}_n(t) = \tilde{X}(t) - \tilde{X}_n(t) + \beta_n t$ ,

$$\sup_{0 \leq t \leq 1} |\tilde{Y}_n(t)| = \sup_{t \in T} |\tilde{Y}_n(t)|.$$

Let  $T_N$  be finite subsets of  $T$ ,  $T_N \uparrow T$ . Then  $\sup_{t \in T} |\tilde{Y}_n(t)| = \limsup_n \sup_{t \in T_N} |\tilde{Y}_n(t)|$ .

Because  $\{\tilde{Y}_n(t)\}$  is a process with stationary, independent increments, we can apply Skorokhod's lemma:

$$P\left(\sup_{t \in T_N} |\tilde{Y}_n(t)| > 2y\right) \leq \frac{1}{1 - C_N} P(|\tilde{Y}_n(1)| > y),$$

where

$$C_N \leq \sup_{0 \leq t \leq 1} P(|\tilde{Y}_n(t)| > y).$$

Let  $\psi_n(u)$  be the exponent function of  $\tilde{Y}_n(t)$ . Apply inequality 8.29 to write:

$$P(|\tilde{Y}_n(t)| > y) \leq \alpha y \int_0^{y^{-1}} |1 - e^{t\psi_n(u)}| du.$$

For  $|e^z| \leq 1$ , there is a constant  $\gamma < \infty$  such that  $|e^z - 1| < \gamma |z|$ . Hence,

$$P(|\tilde{Y}_n(t)| > y) \leq \alpha \gamma y \int_0^{y^{-1}} |\psi_n(u)| du.$$

Since  $\psi_n(u) \xrightarrow{uc} 0$ , then  $C_N \rightarrow 0$ , leading to

$$\sup_{0 \leq t \leq 1} |\tilde{Y}_n(t)| \xrightarrow{P} 0.$$

Thus we can find a subsequence  $\{n'\}$  such that a.s.

$$\tilde{X}_{n'}(t) - \beta_{n'} t \xrightarrow{uc} \tilde{X}(t).$$

Some reflection on this convergence allows the conclusion that the process  $\tilde{X}_{n'}(t)$  can be identified with  $\tilde{X}(D_{n'}^c, t)$ . Therefore  $\{\tilde{X}(t)\}$  and  $\{\tilde{X}_{n'}(t)\}$  have the same joint distribution as  $\{X(t)\}$ ,  $\{X(D_{n'}^c, t)\}$ . Using this fact proves the theorem.

If a function  $x(t) \in D([0, \infty))$  is, at all times  $t$ , the sum of all its jumps up to time  $t$ , then it is called a pure jump function. However, this is not well defined if the sum of the lengths of the jumps is infinite. Then  $x(t)$  is the sum of positive and negative jumps which to some extent cancel each other out to produce  $x(t)$ , and the order of summation of the jumps up to time  $t$  becomes important. We define:

**Definition 14.28.** If there is a sequence of neighborhoods  $D_n \downarrow \{0\}$  such that

$$\sum_{\tau \leq t} \chi_{D_n^c}(x(\tau) - x(\tau-)) \xrightarrow{uc} x(t),$$

then  $x(t) \in D([0, \infty))$  is called a pure jump function.

Many interesting processes of the type studied in this section have the property that there is a sequence of neighborhoods  $D_n \downarrow \{0\}$  such that

$$\beta = \lim_n \int_{D_n^c} \frac{x}{1 + x^2} \mu(dx)$$

exists and is finite. Under this assumption we have

**Corollary 14.29.** Almost every sample path of  $\{X(t) - \beta t\}$ ,  $t \geq 0$ , is a pure jump function.

Another interesting consequence of this construction is the following: Take neighborhoods  $D_n \downarrow \{0\}$  and take  $\{X(D_n^c, t)\}$ , as above, the process of

jumps of size greater than  $D_n$ . Restrict the processes to a finite time interval  $[0, t_0]$  and consider any event  $A$  such that for every  $n$ ,

$$A \in \mathcal{F}(X(t) - X(D_n^c, t), 0 \leq t \leq t_0).$$

Then

**Proposition 14.30.**  $P(A)$  is zero or one.

*Proof.* Let  $B_n = D_{n+1} - D_n$ ; then the processes  $\{X(B_n, t)\}$ ,  $t \in [0, t_0]$  are independent and  $A$  is measurable:

$$\mathcal{F}(\{X(B_n, t)\}, \{X(B_{n+1}, t)\}, \dots, t \in [0, t_0])$$

for every  $n$ . Apply a slight generalization of the Kolmogorov zero-one law to conclude  $P(A) = 0$  or  $1$ .

The results of this section make it easier to understand why infinitely divisible laws were developed to use in the context of processes with independent increments earlier than in the central limit problem. The processes of jumps of different sizes proceed independently of one another, and the jump process of jumps of size  $[x, x + \Delta x)$  contributes a Poisson component with exponent function approximately equal to

$$(e^{iux} - 1)\mu([x, x + \Delta x)).$$

The fact that the measure  $\mu$  governs the number and the size of jumps is further exposed in the following problems, all referring to a process  $\{X(t)\}$ ,  $t \geq 0$  with stationary, independent increments and exponent function having only an integral component.

### Problems

9. Show that for any set  $B \in \mathcal{B}_1$  bounded away from the origin, the process of jumps of size  $B$ ,  $\{X(B, t)\}$ ,  $t \geq 0$  has stationary, independent increments with exponent function

$$\int_B (e^{iux} - 1)\mu(dx),$$

and the processes  $\{X(t) - X(B, t)\}$ ,  $\{X(B, t)\}$  are independent.

10. For  $B$  as above, show that the expected number of jumps of size  $B$  in  $0 \leq t \leq 1$  is  $\mu(B)$ .

11. For  $B$  as above let  $t^*$  be the time of the first jump of size  $B$ . For  $C \subset B$ ,  $C \in \mathcal{B}_1$ , prove that

$$P(X(t^*) - X(t^* -) \in C) = \mu(C)/\mu(B).$$

12. Show that except for a set of probability zero, either all sample functions of  $\{X(t)\}$ ,  $t \in [0, t_0]$  have infinite variation or all have finite variation.



13. Show that for  $\{X(t)\}$ ,  $t \geq 0$  as above, all sample paths have finite variation on every finite time interval  $[0, t_0]$  if and only if

$$\int_{|x| \leq 1} |x| \mu(dx) < \infty.$$

[Take  $t_0 = 1$ . The function

$$V_n = \sum_{k=1}^{2^n} |X(k/2^n) - X(k-1/2^n)|$$

is monotonically nondecreasing. Now compute  $Ee^{-\lambda V_n}$  for  $\lambda > 0$ .]

## 9. EXAMPLES

Consider the first passage time  $\tau_\xi^*$  of a Brownian motion  $X(t)$  to the point  $\xi$ . Denote  $Z(\xi) = \tau_\xi^*$ . For  $\xi_1, \xi_2 > 0$ ,

$$\tau_{\xi_2+\xi_1}^* - \tau_{\xi_1}^* = \inf \{t; X(t + \tau_{\xi_1}^*) - X(\tau_{\xi_1}^*) = \xi_2\}.$$

By the strong Markov property,  $Z(\xi_1 + \xi_2) - Z(\xi_1)$  is independent of  $\mathcal{F}(Z(\xi), \xi \leq \xi_1)$  and is distributed as  $Z(\xi_2)$ . Thus, to completely characterize  $Z(\xi)$ , all we need is its characteristic function. From Chapter 13,

$$e^{iuZ(\xi)} = e^{\xi v(u)},$$

$$v(u) = \begin{cases} -|u|^{1/2} e^{i\pi/4}, & u > 0, \\ -|u|^{1/2} e^{-i\pi/4}, & u < 0. \end{cases}$$

This is the characteristic function of a stable distribution with exponent  $\frac{1}{2}$ . The jump measure  $\mu(dx)$  is given by

$$\mu(dx) = \begin{cases} 0, & x < 0, \\ c \frac{dx}{x^{3/2}}, & x > 0. \end{cases}$$

Doing some definite integrals gives  $c = \frac{1}{2} \sqrt{\pi}$ .

If the characteristic function of a process with stationary, independent increments is stable with exponent  $\alpha$ , call the process *stable with exponent*  $\alpha$ . For  $0 < \alpha < 1$ ,

$$\int_0^\infty (e^{iux} - 1) \frac{dx}{x^{1+\alpha}}$$

exists. The processes with exponent functions

$$v(u) = m \int_0^\infty (e^{iux} - 1) \frac{dx}{x^{1+\alpha}}$$

are the limit of processes  $\{X_n(t)\}$  with exponent functions

$$(14.31) \quad \psi_n(u) = m \int_{1/n}^{\infty} (e^{iux} - 1) \frac{dx}{x^{1+\alpha}}.$$

These latter processes have only upward jumps. Hence all paths of  $\{X(t)\}$  are nondecreasing pure jump functions.

Stable processes of exponent  $\geq 1$  having nondecreasing sample paths do not exist. If  $\mathcal{L}(X(t)) = \mathcal{L}(-X(t))$ , the process is symmetric. Bochner [11] noted that it was possible to construct the symmetric stable processes from Brownian motion by a random time change. Take a sample space with a normalized Brownian motion  $X(t)$  and a stable process  $Z(t)$  defined on it such that  $Z(t)$  has nondecreasing pure jump sample paths and  $\mathcal{F}(X(t), t \geq 0)$ ,  $\mathcal{F}(Z(t), t \geq 0)$  are independent.

**Theorem 14.32.** *If  $Z(t)$  has exponent  $\alpha$ ,  $0 < \alpha < 1$ , then the process  $Y(t) = X(Z(t))$  is a stable symmetric process of exponent  $2\alpha$ .*

*Proof.* The idea can be seen if we write

$$Y(t + \tau) - Y(t) = X((Z(t + \tau) - Z(t)) + Z(t)) - X(Z(t)).$$

Then, given  $Z(t)$ , the process  $Y(t + \tau) - Y(t)$  looks just as if it were the  $Y(\tau)$  process, independent of  $Y(s)$ ,  $s \leq t$ .

For a formal proof, take  $Z_n(t)$  the process of jumps of  $Z(t)$  larger than  $[0, 1/n)$ . Its exponent function is

$$\psi_n(u) = m \int_{1/n}^{\infty} (e^{iux} - 1) \frac{dx}{x^{1+\alpha}}.$$

The jumps of  $Z_n(t)$  occur at the jump times of a Poisson process with intensity

$$\lambda_n = m \int_{1/n}^{\infty} \frac{dx}{x^{1+\alpha}},$$

and the jumps have magnitude  $Y_1, Y_2, \dots$  independent of one another and of the jump times, and are identically distributed. Thus  $X(Z_n(t))$  has jumps only at the jump times of the Poisson process. The size of the  $k$ th jump is

$$U_k = \begin{cases} X(Y_k + \dots + Y_1) - X(Y_{k-1} + \dots + Y_1), & k > 1, \\ X(Y_1), & k = 1. \end{cases}$$

By an argument almost exactly the same as the proof of the strong Markov property for Brownian motion,  $U_k$  is independent of  $U_{k-1}, \dots, U_1$  and has the same distribution as  $U_1$ . Therefore  $X(Z_n(t))$  is a process with stationary, independent increments. Take  $\{n'\}$  so that  $Z_{n'}(t) \xrightarrow{uc} Z(t)$  a.s. Use continuity of

Brownian motion to get  $X(Z_n(t)) \xrightarrow{\text{a.s.}} X(Z(t))$  for every  $t$ . Thus  $X(Z(t))$  is a process with stationary, independent increments. To get its characteristic function, write

$$\begin{aligned} E(e^{iuX(Z(1))} | Z(1) = z) &= Ee^{iuX(z)} \\ &= e^{-zu^2/2}. \end{aligned}$$

Therefore

$$\begin{aligned} Ee^{iuX(Z(1))} &= Ee^{-(u^2/2)Z(1)} \\ &= e^{-c|u|^{2\alpha}}. \end{aligned}$$

## 10. A REMARK ON A GENERAL DECOMPOSITION

Suppose  $\{X(t)\}$ ,  $t \geq 0$  is a process with stationary, independent increments and exponent function

$$\psi(u) = iu\beta - \frac{\sigma^2 u^2}{2} + \int_{(0)^c} \left( e^{iux} - 1 - \frac{iux}{1+x^2} \right) \mu(dx).$$

Since  $iu\beta - \sigma^2 u^2/2$  is the exponent function for a Brownian motion, a natural expectation is that  $X(t) = X^{(1)}(t) + X^{(2)}(t)$ , where  $\{X^{(1)}(t)\}$ ,  $t \geq 0$ , is a Brownian motion with drift  $\beta$  and variance  $\sigma^2$ , and  $\{X^{(2)}(t)\}$ ,  $t \geq 0$ , is a process with stationary, independent increments with exponent function having only an integral component, and that the two processes are independent. This is true, in fact, and can be proved by the methods of Sections 7 and 8. But as processes with stationary, independent increments appear in practice either as a Brownian motion, or a process with no Brownian component, we neglect the proof of this decomposition.

## NOTES

For more material on processes with stationary, independent increments, see Doob [39, Chap. 8] and Paul Lévy's two books [103 and 105]. These latter two are particularly good on giving an intuitive feeling for what these processes look like. Of course, for continuous parameter martingales, the best source is Doob's book. The sample path properties of a continuous parameter martingale were given by Doob in 1951 [38], and applied to processes with independent increments.

Processes with independent increments had been introduced by de Finnett in 1929. Their sample path properties were studied by Lévy in 1934 [102]. He then proved Theorem 14.20 as generalized to processes with independent increments, not necessarily stationary. Most of the subsequent decomposition and building up from Poisson processes follows Lévy also, in particular [103, p. 93]. The article by Ito [75] makes this superposition idea more precise by defining an integral over Poisson processes.

## CHAPTER 15

# MARKOV PROCESSES, INTRODUCTION AND PURE JUMP CASE

### 1. INTRODUCTION AND DEFINITIONS

Markov processes in continuous time are, as far as definitions go, a straightforward extension of the Markov dependence idea.

**Definition 15.1.** A process  $\{X(t)\}$ ,  $t \geq 0$  is called Markov with state space  $F \in \mathcal{B}_1$ , if  $X(t) \in F$ ,  $t \geq 0$ , and for any  $B \in \mathcal{B}_1(F)$ ,  $t, \tau \geq 0$ .

$$(15.2) \quad P(X(t + \tau) \in B \mid X(s), s \leq t) = P(X(t + \tau) \in B \mid X(t)) \quad \text{a.s.}$$

To verify that a process is Markov, all we need is to have for any  $t_{n+1} \geq t_n \geq \dots \geq t_1 \geq 0$ ,

$$P(X(t_{n+1}) \in B \mid X(t_n), \dots, X(t_1)) = P(X(t_{n+1}) \in B \mid X(t_n)) \quad \text{a.s.}$$

Since finite dimensional sets determine  $\mathcal{F}(X(s), s \leq t)$ , this extends to (15.2).

The Markov property is a statement about the conditional probability at the one instant  $t + \tau$  in the future. But it extends to a general statement about the future, given the present and past:

**Proposition 15.3.** If  $\{X(t)\}$  is Markov, then for  $A \in \mathcal{F}(X(\tau), \tau \geq t)$ ,

$$P(A \mid X(s), s \leq t) = P(A \mid X(t)) \quad \text{a.s.}$$

The proof of this is left to the reader.

**Definition 15.4.** By Theorem 4.30, for every  $t_2 > t_1$  a version  $p_{t_2, t_1}(B \mid x)$  of  $P(X(t_2) \in B \mid X(t_1) = x)$  can be selected such that  $p_{t_2, t_1}(B \mid x)$  is a probability on  $\mathcal{B}_1(F)$  for  $x$  fixed, and  $\mathcal{B}_1(F)$  measurable in  $x$  for  $B$  fixed. Call these a set of transition probabilities for the process.

**Definition 15.5.**  $\pi(\cdot)$  on  $\mathcal{B}_1(F)$  given by  $\pi(B) = P(X(0) \in B)$  is the initial distribution for the process.

The importance of transition probabilities is that the distribution of the process is completely determined by them if the initial distribution is specified.

This follows from

**Proposition 15.6.** For  $B_n, \dots, B_0 \in \mathcal{B}_1(F)$ ,  $t_n > t_{n-1} > \dots > t_0 = 0$ ,

$$(15.7) \quad P(X(t_n) \in B_n, \dots, X(t_1) \in B_1, X(t_0) \in B_0) \\ = \int_{B_n} \dots \int_{B_1} \int_{B_0} p_{t_n, t_{n-1}}(dx_n | x_{n-1}) \dots p_{t_1, t_0}(dx_1 | x_0) \pi(dx_0).$$

The proof is the same as for 7.3.

A special case of (15.7) are the Chapman-Kolmogorov equations. Reason this way: To get from  $x$  at time  $\tau$  to  $B$  at time  $t > \tau$ , fix any intermediate time  $s$ ,  $t > s > \tau$ . Then the probability of all the paths that go from  $x$  to  $B$  through the small neighborhood  $dy$  of  $y$  at time  $s$  is given (approximately) by  $p_{t,s}(B | y)p_{s,\tau}(dy | x)$ . So summing over  $dy$  gets us to the Chapman-Kolmogorov equations,

$$(15.8) \quad p_{t,\tau}(B | x) = \int p_{t,s}(B | y)p_{s,\tau}(dy | x).$$

Actually, what is true is

**Proposition 15.9.** The equation (15.8) holds a.s. with respect to  $P(X(\tau) \in dx)$ .

*Proof.* From (15.7), for every  $C \in \mathcal{B}_1(F)$ ,

$$P(X(t) \in B, X(\tau) \in C) = \iint_C p_{t,s}(B | y)p_{s,\tau}(dy | x)P(X(\tau) \in dx).$$

Taking the Radon derivative with respect to  $P(X(\tau) \in dx)$  gives the result.

One rarely starts with a process on a sample space  $(\Omega, \mathcal{F}, P)$ . Instead, consistent distribution functions are specified, and then the process constructed. For a Markov process, what gets specified are the transition probabilities and the initial distribution. Here there is a divergence from the discrete time situation in which the one-step transition probabilities  $P(X_{n+1} \in B | X_n = x)$  determine all the multiple-step probabilities. There are no corresponding one-step probabilities here; the probabilities  $\{p_{t,\tau}(B | x)\}$ ,  $t > \tau \geq 0$  must all be specified, and they must satisfy among themselves at least the functional relationship (15.8).

## 2. REGULAR TRANSITION PROBABILITIES

**Definition 15.10.** A set of functions  $\{p_{t,\tau}(B | x)\}$  defined for all  $t > \tau \geq 0$ ,  $B \in \mathcal{B}_1(F)$ ,  $x \in F$  is called a regular set of transition probabilities if

- 1)  $p_{t,\tau}(B | x)$  is a probability on  $\mathcal{B}_1(F)$  for  $x$  fixed and  $\mathcal{B}_1(F)$  measurable in  $x$  for  $B$  fixed,

2) for every  $B \in \mathcal{B}_1(F)$ ,  $x \in F$ ,  $t > s > \tau \geq 0$ ,

$$p_{t,\tau}(B | x) = \int p_{t,s}(B | y) p_{s,\tau}(dy | x).$$

To be regular, then, a set of transition probabilities must satisfy the Chapman-Kolmogorov equations identically.

**Theorem 15.11.** *Given a regular set of transition probabilities and a distribution  $\pi(dx_0)$ , define probabilities on cylinder sets by: for  $t_n > t_{n-1} > \cdots > t_0 = 0$ ,  $B_n, \dots, B_0 \in \mathcal{B}_1(F)$ ,*

$$(15.12) \quad P(X(t_n) \in B_n, \dots, X(t_0) \in B_0) \\ = \int_{B_n} \cdots \int_{B_0} p_{t_n, t_{n-1}}(dx_n | x_{n-1}) \cdots p_{t_1, t_0}(dx_1 | x_0) \pi(dx_0).$$

*These are consistent and the resultant process  $\{X(t)\}$  is Markov with the given functions  $p_{t,\tau}$ ,  $\pi$  as transition probabilities and initial distribution.*

*Proof.* The first verification is consistency. Let  $B_k = F$ . The expression in  $P(X(t_n) \in B_n, \dots)$  that involves  $x_k, t_k$  is an integration with respect to the probability defined for  $B \in \mathcal{B}_1(F)$  by

$$\int_F p_{t_{k+1}, t_k}(B | x_k) p_{t_k, t_{k-1}}(dx_k | x_{k-1}).$$

By the Chapman-Kolmogorov equations, this is exactly

$$p_{t_{k+1}, t_{k-1}}(B | x_{k-1}),$$

which eliminates  $t_k, x_k$  and gives consistency. Thus, a bit surprisingly, the functional relations (15.8) are the key to consistency. Now extend and get a process  $\{X(t)\}$  on  $(F^{[0, \infty)}, \mathcal{B}^{[0, \infty)}(F))$ . To verify the remainder, let  $A \in \mathcal{B}_n(F)$ . By extension from the definition,

$$P(X(t_n) \in B, (X(t_{n-1}), \dots, X(t_0)) \in A) \\ = \int_{(x_{n-1}, \dots, x_0) \in A} p_{t_n, t_{n-1}}(B | x_{n-1}) P(X(t_{n-1}) \in dx_{n-1}, \dots, X(t_0) \in dx_0).$$

Now, take the Radon derivative to get the result.

One result of this theorem is that there are no functional relationships other than those following from the Chapman-Kolmogorov equations that transition probabilities must satisfy in general.

Another convention we make starting from a regular set of transition probabilities is this: Let  $P_{(x,\tau)}(\cdot)$  be the probability on the space of paths  $F^{[0, \infty)}$ , gotten by "starting the process out at the point  $x$  at time  $\tau$ ." More specifically, let  $P_{(x,\tau)}(\cdot)$  be the probability on  $\mathcal{B}^{[0, \infty)}(F)$  extended from (15.12), where  $t_0 = \tau$  and  $\pi(dx_0)$  concentrated on the point  $\{x\}$  are used. So  $P_{(x,\tau)}(\cdot)$  is well-defined for all  $x, \tau$  in terms of the transition probabilities only.

**Convention.** For  $C \in \mathcal{F}(X(s), s \leq \tau)$ ,  $A \in \mathcal{B}^{(0, \infty)}(F)$  always use the version of  $P(X(\tau + \cdot) \in A \mid X(\tau) = x, C)$  given by  $P_{(x, \tau)}(A)$ .

Accordingly, we use the transition probabilities not only to get the distribution of the process but also, to manufacture versions of all important conditional probabilities. The point of requiring the Chapman-Kolmogorov equations to hold identically rather than a.s. is that if there is an  $x, B$ ,  $t > s > \tau \geq 0$ , such that

$$p_{t, \tau}(B \mid x) \neq \int p_{t, s}(B \mid y) p_{s, \tau}(dy \mid x),$$

then these transition probabilities can not be used to construct “the process starting from  $x, \tau$ .”

Now, because we wish to study the nature of a Markov process as governed by its transition probabilities, no matter what the initial distribution, we enlarge our nomenclature. Throughout this and the next chapter when we refer to a Markov process  $\{X(t)\}$  this will no longer refer to a single process. Instead, it will denote the totality of processes having the same transition probabilities but with all possible different initial starting points  $x$  at time zero. However, we will use only coordinate representation processes, so the measurability of various functions and sets will not depend on the choice of a starting point or initial distribution for the process.

### 3. STATIONARY TRANSITION PROBABILITIES

**Definition 15.13.** Let  $\{p_{t, \tau}\}$  be a regular set of transition probabilities. They are called stationary if for all  $t > \tau \geq 0$ ,

$$p_{t, \tau}(B \mid x) \equiv p_{t-\tau, 0}(B \mid x).$$

In this case, the  $p_t(B \mid x) = p_{t, 0}(B \mid x)$  are referred to as the transition probabilities for the process.

Some simplification results when the transition probabilities are stationary. The Chapman-Kolmogorov equations become

$$(15.14) \quad p_{t+\tau}(B \mid x) = \int p_t(B \mid y) p_{\tau}(dy \mid x).$$

For any  $A \in \mathcal{B}^{(0, \infty)}(F)$ ,  $P_{(x, \tau)}(A) = P_{(x, 0)}(A)$ ; the probabilities on the path-space of the process are the same no matter when the process is started. Denote for any  $A \in \mathcal{B}^{(0, \infty)}(F)$ ,  $f(\cdot)$  on  $F^{(0, \infty)}$  measurable  $\mathcal{B}^{(0, \infty)}(F)$ ,

$$(15.15) \quad P_x(A) = P_{(x, 0)}(A),$$

$$E_x f(\cdot) = \int f(y) P_x(dy).$$

Assume also from now on that for any initial distribution,  $X(t) \xrightarrow{P} X(0)$  as  $t \rightarrow 0$ . This is equivalent to the statement,

**Definition 15.16.** Transition probabilities  $p_t(B | x)$  are called standard if

$$p_t(\cdot | x) \xrightarrow{D} \delta_{\{x\}}(\cdot),$$

for all  $x \in F$ , where  $\delta_{\{x\}}(\cdot)$  denotes the distribution with unit mass on  $\{x\}$ .

There is another property that will be important in the sequel. Suppose we have a stopping time  $\tau^*$  for the process  $X(t)$ . The analog of the restarting property of Brownian motion is that the process  $X(t + \tau^*)$ , given everything that happened up to time  $\tau^*$ , has the same distribution as the process  $X(t)$  starting from the point  $X(\tau^*)$ .

**Definition 15.17.** A Markov process  $\{X(t)\}$  with stationary transition probabilities is called strong Markov if for every stopping time  $\tau^*$  [see (12.40)], every starting point  $x \in F$ , and set  $A \in \mathcal{B}^{[0, \infty)}(F)$ ,

$$P_x(X(\cdot + \tau^*) \in A | X(s), s \leq \tau^*) = P_{X(\tau^*)}(X(\cdot) \in A) \quad \text{a.s. } P_x.$$

Henceforth, call a stopping time for a Markov process a *Markov time*.

It's fairly clear from the definitions that for fixed  $\tau > 0$ ,

$$(15.18) \quad P_x(X(\cdot + \tau) \in A | X(\tau)) = P_{X(\tau)}(X(\cdot) \in A) \quad \text{a.s. } P_x.$$

In fact,

**Proposition 15.19.** If  $\tau^*$  assumes at most a countable number of values  $\{\tau_k\}$ , then

$$P_x(X(\cdot + \tau^*) \in A | X(s), s \leq \tau^*) = P_{X(\tau^*)}(X(\cdot) \in A) \quad \text{a.s. } P_x.$$

*Proof.* Take  $C \in \mathcal{F}(X(s), s \leq \tau^*)$ . Let  $\varphi(x)$  be any bounded measurable function,  $\tilde{\varphi}(x) = E_x \varphi(X(t))$ . We prove the proposition first for  $A$  one-dimensional. Take  $\varphi$  the set indicator of  $A$ , then

$$(15.20) \quad \int_C \varphi(X(t + \tau^*)) dP_x = \sum_k \int_{C \cap \{\tau^* = \tau_k\}} \varphi(X(t + \tau_k)) dP_x.$$

By definition,  $C \cap \{\tau^* = \tau_k\} \in \mathcal{F}(X(s), s \leq \tau_k)$ , so

$$\begin{aligned} \int_{C \cap \{\tau^* = \tau_k\}} \varphi(X(t + \tau_k)) dP_x &= \int_{C \cap \{\tau^* = \tau_k\}} E[\varphi(X(t + \tau_k)) | X(s), s \leq \tau_k] dP_x \\ &= \int_{C \cap \{\tau^* = \tau_k\}} \tilde{\varphi}(X(\tau_k)) dP_x, \end{aligned}$$

which does it. The same thing goes for  $\varphi$ , a measurable function of many variables, and then for the general case.

It is not unreasonable to hope that the strong Markov property would hold in general. It doesn't! But we defer an example until the next chapter.



One class of examples of strong Markov processes with standard stationary transition probabilities are the processes with stationary, independent increments, where the sample paths are taken right-continuous. Let  $X(t)$  be such a process. If at time  $t$  the particle is at the point  $x$ , then the distribution at time  $t + \tau$  is gotten by adding to  $x$  an increment independent of the path up to  $x$  and having the same distribution as  $X(\tau)$ . Thus, these processes are Markov.

### Problems

1. For  $X(t)$ , a process with stationary, independent increments, show that it is Markov with one set of transition probabilities given by

$$p_t(B | x) = P(X(t) \in B - x),$$

and that this set is regular and standard.

2. Show that processes with stationary, independent increments and right-continuous sample paths are strong Markov.

3. Show that the functions  $P_x(A)$ ,  $E_x f(\cdot)$  of (15.15) are  $\mathcal{B}_1(F)$ -measurable.

4. Show that any Markov process with standard stationary transition probabilities is continuous from the right in probability.

### 4. INFINITESIMAL CONDITIONS

What do Markov processes look like? Actually, what do their sample paths and transition probabilities look like? This problem is essentially one of connecting up global behavior with local behavior. Note, for example, that if the transition probabilities  $p_t$  are known for all  $t$  in any neighborhood of the origin, then they are determined for all  $t > 0$  by the Chapman-Kolmogorov equations. Hence, one suspects that  $p_t$  would be determined for all  $t$  by specifying the limiting behavior of  $p_t$  as  $t \rightarrow 0$ . But, then, the sample behavior will be very immediately connected with the behavior of  $p_t$  near  $t = 0$ .

To get a feeling for this, look at the processes with stationary, independent increments. If it is specified that

$$P_x(|X(t) - x| > \epsilon) = o(t),$$

then the process is Brownian motion, all the transition probabilities are determined, and all sample paths are continuous. Conversely, if all sample paths are given continuous, the above limiting condition at  $t = 0$  must hold.

At the other end, suppose one asks for a process  $X(t)$  with stationary, independent increments having all sample paths constant except for isolated jumps. Then (see Section 6, Chapter 14) the probability of no jump in the

time interval  $[0, t]$  is given by  $e^{-\lambda t}$ , so

$$P_0(X(t) = 0) = 1 - \lambda t + o(t).$$

If there is a jump, with magnitude governed by  $F(x)$ , then for  $B \in \mathcal{B}_1$ ,  $\{0\} \notin B$ ,

$$P_0(X(t) \in B) = \lambda t F(B) + o(t).$$

Conversely, if there is a process  $X(t)$  with stationary, independent increments, and a  $\lambda, F$  such that the above conditions hold as  $t \rightarrow 0$ , it is easy to check that the process must be of the jump type with exponent function given by

$$\psi(u) = \lambda \int_{\{0\}^c} (e^{iux} - 1) F(dx).$$

In general now, let  $\{X(t)\}$  be any Markov process with stationary transition probabilities. Take  $f(x)$  a bounded  $\mathcal{B}_1$ -measurable function on  $F$ . Consider the class of these functions such that the limit as  $t \downarrow 0$  of

$$(15.21) \quad \frac{E_x f(X(t)) - f(x)}{t}$$

exists for all  $x \in F$ . Denote the resulting function by  $(Sf)(x)$ .  $S$  is called the *infinitesimal operator* and summarizes the behavior of the transition probabilities as  $t \rightarrow 0$ . The class of bounded measurable functions such that the limit in (15.21) exists for all  $x \in F$ , we will call the domain of  $S$ , denoted by  $\mathcal{D}(S)$ . For example, for Poisson-like processes,

$$\begin{aligned} E_x f(X(t)) &= E_0 f(X(t) + x) \\ &= \lambda t \int_{\{0\}^c} f(y + x) F(dy) + f(x)[1 - \lambda t] + o(t). \end{aligned}$$

Define a measure  $\mu(B; x)$  by

$$\mu(B; x) = \lambda F(B - x), \quad x \notin B, \quad \text{and} \quad \mu(\{x\}; x) = 0.$$

Then  $S$  for this process can be written as

$$(Sf)(x) = \int (f(y) - f(x)) \mu(dy; x).$$

In this example, no further restrictions on  $f$  were needed to make the limit as  $t \rightarrow 0$  exist. Thus  $\mathcal{D}(S)$  consists of all bounded measurable functions.

For Brownian motion, take  $f(x)$  continuous and with a continuous, bounded second derivative. Write

$$f(y) = f(x) + (y - x)f'(x) + \frac{(y - x)^2}{2} (f''(x) + \delta_x(y - x)),$$

where

$$\delta_x(h) \leq \sup_{|\xi| \leq |h|} |f''(x + \xi) - f''(x)|.$$

From this,

$$\begin{aligned} E_x f(X(t)) &= E_0 f(X(t) + x) \\ &= f(x) + \frac{1}{2}(E_0 X^2(t))f''(x) + R(t). \end{aligned}$$

Use

$$EX^2(t) = t, \quad R(t) \leq EX^2(t)\delta_x(X(t)) = o(t),$$

to get

$$(Sf)(x) = \frac{1}{2} \frac{d^2}{dx^2} f(x).$$

In this case it is not clear what  $\mathcal{D}(S)$  is, but it is certainly not the set of all bounded measurable functions. These two examples will be typical in this sense: The jumps in the sample paths contribute an integral operator component to  $S$ ; the continuous nonconstant parts of the paths contribute a differential operator component.

Once the behavior of  $p_t$  near  $t = 0$  is specified by specifying  $S$ , the problem of computing the transition probabilities for all  $t > 0$  is present.  $S$  hooks into the transition probabilities in two ways. In the first method, we let the initial position be perturbed. That is, given  $X(0) = x$ , we let a small time  $\tau$  elapse and then condition on  $X(\tau)$ . This leads to the *backwards equations*. In the second method, we perturb on the final position. We compute the distribution up to time  $t$  and then let a small time  $\tau$  elapse. Figures 15.1 and 15.2 illustrate computing  $P_x(X(t + \tau) \in B)$ .

### Backwards Equations

$$P(X(t + \tau) \in B \mid X(0) = x) = E[P(X(t + \tau) \in B \mid X(\tau)) \mid X(0) = x].$$

Letting  $\varphi_t(x) = p_t(B \mid x)$ , we can write the above as

$$\varphi_{t+\tau}(x) = E_x \varphi_t(X(\tau)) \quad \text{or} \quad \varphi_{t+\tau}(x) - \varphi_t(x) = E_x \varphi_t(X(\tau)) - \varphi_t(x).$$

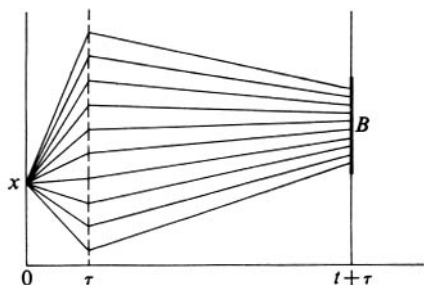


Fig. 15.1 Backwards equations.

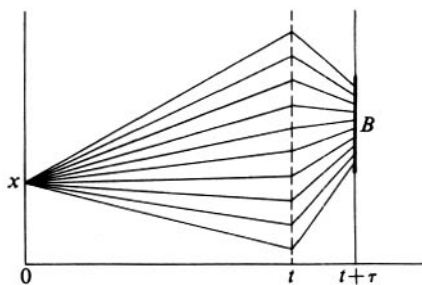


Fig. 15.2 Forwards equations.

Dividing by  $\tau$ , letting  $\tau \rightarrow 0$ , if  $p_t(B | x)$  is smooth enough in  $t, x$ , we find that

$$(15.22) \quad \frac{\partial}{\partial t} p_t(B | x) = (S p_t(B | \cdot))(x),$$

that is, for Brownian motion, the backwards equations are

$$\frac{\partial}{\partial t} p_t(B | x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} p_t(B | x).$$

For the Poisson-like processes,

$$\frac{\partial}{\partial t} p_t(B | x) = \int (p_t(B | y) - p_t(B | x)) \mu(dy; x).$$

**Forwards Equations.** For any  $f$  for which  $Sf$  exists,

$$E_x f(X(t + \tau)) = E_x [E_x(f(X(t + \tau)) | X(t))]$$

or

$$\int f(y) p_{t+\tau}(dy | x) = \int \left[ \int f(z) p_\tau(dz | y) \right] p_t(dy | x).$$

Subtract  $E_x f(X(t))$  from both sides, divide by  $\tau$ , and let  $\tau \downarrow 0$ . With enough smoothness,

$$\int f(y) \frac{\partial}{\partial t} p_t(dy | x) = \int (Sf)(y) p_t(dy | x).$$

Thus, if  $S$  has an adjoint  $S^*$ , the equations are

$$(15.23) \quad \frac{\partial}{\partial t} p_t(B | x) = (S^* p_t(\cdot | x))(B).$$

For Poisson-like processes,

$$\int (Sf)(y) p_t(dy | x) = \iint (f(z) - f(y)) \mu(dz; y) p_t(dy | x)$$

so the forwards equations are

$$\frac{\partial}{\partial t} p_t(B | x) = \int \bar{\mu}(B; y) p_t(dy | x),$$

where  $\bar{\mu}(B; y) = \mu(B; y)$ ,  $y \notin B$ , and  $\bar{\mu}(\{y\}; y) = -\mu(F; y)$ . If  $p_t(dy | x) \ll \mu(dy)$  for all  $t, x$ , then take  $S^*$  to be the adjoint with respect to  $\mu(dy)$ . For example, for Brownian motion  $p_t(dy | x) \ll dy$ . For all  $f(y)$  with continuous second derivatives vanishing off finite intervals,

$$\int \frac{1}{2} f''(y) p_t(y | x) dy = \int f(y) \left( \frac{1}{2} \frac{\partial^2}{\partial y^2} p_t(y | x) \right) dy,$$

where  $p_t(y | x)$  denotes (badly) the density of  $p_t(dy | x)$  with respect to  $dy$ . Hence the forwards equation is

$$\frac{\partial}{\partial t} p_t(y | x) = \frac{1}{2} \frac{\partial^2}{\partial y^2} p_t(y | x).$$

The forwards or backwards equations, together with the boundary condition  $p_t(\cdot | x) \xrightarrow{D} \delta_{(x)}(\cdot)$  as  $t \rightarrow 0$  can provide an effective method of computing the transition probabilities, given the infinitesimal conditions. But the questions regarding the existence and uniqueness of solutions are difficult to cope with. It is possible to look at these equations analytically, forgetting their probabilistic origin, and investigate their solutions. But the most illuminating approach is a direct construction of the required processes.

## 5. PURE JUMP PROCESSES

**Definition 15.24.** A Markov process  $\{X(t)\}$  will be called a pure jump process if, starting from any point  $x \in F$ , the process has all sample paths constant except for isolated jumps, and right-continuous.

**Proposition 15.25.** If  $\{X(t)\}$  is a pure jump process, then it is strong Markov.

*Proof.* By 15.19, for  $\varphi$  bounded and measurable  $\mathcal{B}_k$ , and

$$E_x \varphi(X(t_k), \dots, X(t_0)) = \bar{\varphi}(x),$$

we have for any  $C \in \mathcal{F}(X(t), t \leq t_n^*)$ , where  $t_n^*$  takes on only a countable number of values and  $t_n^* \downarrow t^*$ .

$$\int_C \varphi(X(t_k + t_n^*), \dots, X(t_0 + t_n^*)) dP_x = \int_C \bar{\varphi}(X(t_n^*)) dP_x.$$

Since  $t^* \leq t_n^*$ , this holds for all  $C \in \mathcal{F}(X(t), t \leq t^*)$ . Since all paths are constant except for jumps and right-continuous,

$$\bar{\varphi}(X(t_n^*)) \rightarrow \bar{\varphi}(X(t^*)),$$

$$\varphi(X(t_k + t_n^*), \dots) \rightarrow \varphi(X(t_k + t^*), \dots)$$

for every sample path. Taking limits in the above integral equality proves the proposition.

Assume until further notice that  $\{X(t)\}$  is a pure jump process.

**Definition 15.26.** Define the time  $T$  of the first jump as

$$T = \inf\{t; X(t) \neq X(0)\}.$$

**Proposition 15.27.**  $T$  is a Markov time.

with  $0 \leq \lambda(x) < \infty$ . If  $\lambda(x) = 0$ , then the state  $x$  is absorbing, that is,  $P_x(X(t) = x) = 1$ , for all  $t$ . The measurability of  $\lambda(x)$  follows from the measurability of  $P_x(T > t)$ .

**Corollary 15.29.** *For a pure jump process, let*

$$P_x(T > t) = e^{-\lambda(x)t}, \quad P_x(X(T) \in B) = p(B; x).$$

*Then at every point  $x$  in  $F$ ,*

$$\begin{aligned} p_t(\{x\} | x) &= 1 - \lambda(x)t + o(t), \\ p_t(B | x) &= (\lambda(x)t + o(t))p(B; x), \quad x \notin B. \end{aligned}$$

*Proof.* Suppose that we prove that the probability of two jumps in time  $t$  is  $o(t)$ ; then both statements are clear because

$$\begin{aligned} P_x(X(t) = x) &= P_x(X(t) = x, \text{ no more than one jump}) + o(t) \\ &= P_x(T > t) + o(t) \\ &= 1 - \lambda(x)t + o(t). \end{aligned}$$

Similarly,

$$\begin{aligned} P_x(X(t) \in B) &= P_x(X(t) \in B, \text{ no more than one jump}) + o(t) \\ &= P_x(T \leq t, X(T) \in B) + o(t) \\ &= (\lambda(x)t + o(t))p(B; x). \end{aligned}$$

**Remark.** The reason for writing  $o(t)$  in front of  $p(B; x)$  is to emphasize that  $o(t)/t \rightarrow 0$  uniformly in  $B$ .

We finish by

**Proposition 15.30.** *Let  $T_0$  be the time of the first jump,  $T_0 + T_1$  the time of the second jump. Then*

$$P_x(T_0 + T_1 \leq t) = o(t).$$

*Proof.*  $P_x(T_0 + T_1 \leq t) \leq P_x(T_0 \leq t, T_1 \leq t)$ . Now,

$$P_x(T_1 \leq t | T_0) = E_x[P_x(T_1 \leq t | T_0, X(T_0)) | T_0],$$

so

$$P_x(T_1 \leq t | T_0) = \int (1 - e^{-\lambda(y)t})p(dy; x).$$

This goes to zero as  $t \rightarrow 0$ , by the bounded convergence theorem. The following inequality

$$P_x(T_0 + T_1 \leq t) \leq (1 - e^{-\lambda(x)t}) \cdot \int (1 - e^{-\lambda(y)t})p(dy; x),$$

now proves the stated result.

*Proof.* Let  $T_n = \inf\{k/2^n; X(k/2^n) \neq X(0)\}$ . For every  $\omega$ ,  $T_n \downarrow T$ . Further,

$$\{T_n \leq t\} = \bigcup_{k/2^n \leq t} \{X(k/2^n) \neq X(0)\} \in \mathcal{F}(X(s), s \leq t).$$

The sets  $\{T_n \leq t\}$  are monotonic increasing, and their limit is therefore in  $\mathcal{F}(X(s), s \leq t)$ . If  $t$  is a binary rational, the limit is  $\{T \leq t\}$ . If not, the limit is  $\{T < t\}$ . But,

$$\{T \leq t\} = \{T < t\} \cup \{X(t) \neq X(0)\}.$$

The basic structure of Markov jump processes is given by

**Theorem 15.28.** Under  $P_x$ ,  $T$  and  $X(T)$  are independent and there is a  $\mathcal{B}_1$ -measurable nonnegative function  $\lambda(x)$  on  $F$  such that

$$P_x(T > t) = e^{-\lambda(x)t}.$$

*Proof.* Let  $t + T_1$  be the first exit time from state  $x$  past time  $t$ ; that is,

$$t + T_1 = \inf\{\tau; X(\tau) \neq x, \tau > t\}.$$

Then, for  $x \notin B$ ,

$$\begin{aligned} P_x(X(T) \in B, T > t) &= P_x(X(T) \in B, T > t, X(t) = x) \\ &= P_x(X(t + T_1) \in B, T > t, X(t) = x) \\ &= P_x(X(t + T_1) \in B \mid T > t, X(t) = x) \\ &\quad \cdot P_x(T > t, X(t) = x) \end{aligned}$$

if  $P_x(T > t) > 0$ . Assume this for the moment. Then,

$$\{T > t\} = \{T \leq t\}^c \in \mathcal{F}(X(s), s \leq t),$$

$$\begin{aligned} P_x(X(t + T_1) \in B \mid T > t, X(t) = x) &= P_x(X(t + T_1) \in B \mid X(t) = x) \\ &= P_x(X(T) \in B). \end{aligned}$$

Going back, we have

$$P_x(X(T) \in B, T > t) = P_x(X(T) \in B)P_x(T > t).$$

There must be some  $t_0 > 0$ , such that  $P_x(T > t_0) > 0$ . Take  $0 < t \leq t_0$ ; then

$$\begin{aligned} P_x(T > t + \tau) &= P_x(T > t + \tau, T > t, X(t) = x) \\ &= P_x(T_1 > \tau \mid X(t) = x, T > t)P_x(T > t) \\ &= P_x(T > \tau)P_x(T > t). \end{aligned}$$

Therefore  $\varphi(t) = P_x(T > t)$  is a monotonic nonincreasing function satisfying  $\varphi(t + \tau) = \varphi(t)\varphi(\tau)$  for  $t \leq t_0$ . This implies that there is a parameter  $\lambda(x)$  such that

$$P_x(T > t) = e^{-\lambda(x)t},$$

Define a measure  $\mu(dy; x)$  by

$$\mu(B; x) = \lambda(x)p(B; x), \quad \mu(\{x\}, x) = 0.$$

The infinitesimal operator  $S$  is given, following 15.30, by

$$(15.31) \quad (Sf)(x) = \int (f(y) - f(x))\mu(dy; x),$$

and  $\mathcal{D}(S)$  consists of all bounded  $\mathcal{B}_1(F)$  measurable functions. The following important result holds for jump processes.

**Theorem 15.32.**  $p_t(B | x)$  satisfies the backwards equations.

*Proof.* First, we derive the equation

$$(15.33) \quad p_t(B | x) = \chi_B(x)e^{-\lambda(x)t} + \int_0^t \int_{\{x\}^c} p_{t-\tau}(B | y)\mu(dy; x)e^{-\lambda(x)\tau} d\tau.$$

The intuitive idea behind (15.33) is simply to condition on the time and position of the first jump. To see this, write

$$P_x(X(t) \in B) = P_x(X(t) \in B, T > t) + P_x(X(t) \in B, T \leq t).$$

The first term is  $\chi_B(x)e^{-\lambda(x)t}$ . Reason that to evaluate the second term, if  $T \in d\tau$ , and  $X(T) \in dy$ , then the particle has to get from  $y$  to  $B$  in time  $t - \tau$ . Hence the second term should be

$$\int_0^t \int p_{t-\tau}(B | y)P_x(X(T) \in dy)P_x(T \in d\tau),$$

and this is exactly the second term in (15.33). A rigorous derivation could be given along the lines of the proof of the strong Markov property. But it is easier to use a method involving Laplace transforms which has wide applicability when random times are involved. We sketch this method: First note that since  $X(t)$  is jointly measurable in  $t, \omega$ ,  $p_t(B | x) = E_x \chi_B(X(t))$  is measurable in  $t$ . Define

$$\varphi_s(x) = \int_0^\infty e^{-st} p_t(B | x) dt = E_x \left[ \int_0^\infty e^{-st} \chi_B(X(t)) dt \right], \quad s > 0.$$

Write

$$\int_0^\infty e^{-st} \chi_B(X(t)) dt = \int_0^T e^{-st} \chi_B(X(t)) dt + \int_T^\infty e^{-st} \chi_B(X(t)) dt.$$

The first term is

$$\chi_B(X(0)) \left( \frac{1 - e^{-sT}}{s} \right).$$



The second is

$$e^{-s\tau} \int_0^\infty e^{-st} \chi_B(X(t + \tau)) dt.$$

By the strong Markov property,

$$E_x \left( e^{-s\tau} \int_0^\infty e^{-st} \chi_B(X(t + \tau)) dt \mid X(\tau), \tau \leq \tau \right) = e^{-s\tau} \varphi_s(X(\tau)).$$

Hence

$$\varphi_s(x) = \chi_B(x) \cdot \frac{1}{s + \lambda(x)} + \frac{\lambda(x)}{s + \lambda(x)} \int \varphi_s(y) p(dy; x).$$

This is exactly the transform of (15.33), and invoking the uniqueness theorem for Laplace transforms (see [140]) gets (15.33) almost everywhere ( $dt$ ). For  $\{X(t)\}$  a pure jump process, writing  $p_t(B \mid x)$  as  $E_x \chi_B(X(t))$  makes it clear that  $p_t(B \mid x)$  is right-continuous. The right side of (15.33) is obviously continuous in time; hence (15.33) holds identically.

Multiply (15.33) by  $e^{\lambda(x)t}$  and substitute  $t - \tau = \tau'$  in the second term:

$$e^{\lambda(x)t} p_t(B \mid x) = \chi_B(x) + \int_0^t \int_{\{x\}^c} e^{\lambda(x)\tau} p_\tau(B \mid y) \mu(dy; x) d\tau.$$

Hence  $p_t(B \mid x)$  is differentiable, and

$$\frac{\partial}{\partial t} (e^{\lambda(x)t} p_t(B \mid x)) = e^{\lambda(x)t} \int_{\{x\}^c} p_t(B \mid y) \mu(dy; x).$$

An easy simplification gives the backwards equations.

The forwards equations are also satisfied. (See Chung [16, pp. 224 ff.], for instance.) In fact, most of the questions regarding the transition probabilities can be answered by using the representation

$$(15.34) \quad \chi_B(X(t)) = \sum_{n=0}^{\infty} \chi_{[R_n, R_{n+1})}(t) \chi_B(X(R_n)),$$

where  $R_n$  is the time of the  $n$ th jump,  $R_0 \equiv 0$ , so

$$R_n = T_0 + \cdots + T_{n-1}, \quad n > 0,$$

where the  $T_k$  are the first exit times after the  $k$ th jump. We make use of this in the next section to prove a uniqueness result for the backward equation.

**Problem 5.** Show that a pure jump process  $\{X(t)\}$ ,  $t \geq 0$  is jointly measurable in  $(t, \omega)$  with respect to  $\mathcal{B}_1([0, \infty)) \times \mathcal{F}$ .

## 6. CONSTRUCTION OF JUMP PROCESSES

In modeling Markov processes what is done, usually, is to prescribe infinitesimal conditions. For example: Let  $F$  be the integers, then a population model with constant birth and death rates would be constructed by

specifying that in a small time  $\Delta t$ , if the present population size is  $j$ , the probability of increasing by one is  $r_B j \Delta t$ , where  $r_B$  is the birth rate. The probability of a decrease is  $r_D j \Delta t$  where  $r_D$  is the death rate, and the probability of no change is  $1 - r_D j \Delta t - r_B j \Delta t$ . What this translates into is

$$\begin{aligned} p_t(j+1|j) &= r_B j t + o(t), \quad k \neq j, \\ p_t(j-1|j) &= r_D j t + o(t), \\ p_t(j|j) &= 1 - (r_D + r_B) j t + o(t), \\ p_t(k|j) &= o(t), \quad \text{otherwise.} \end{aligned}$$

In general, countable state processes are modeled by specifying  $q(k|j) \geq 0$ ,  $q(j) = \sum_{k \neq j} q(k|j) < \infty$ , such that

$$\begin{aligned} p_t(k|j) &= q(k|j)t + o(t), \quad k \neq j, \\ p_t(j|j) &= 1 - q(j)t + o(t). \end{aligned}$$

General jump processes are modeled by specifying finite measures  $\mu(B; x)$ , measurable in  $x$  for every  $B \in \mathcal{B}_1(F)$ , such that for every  $x$ ,

$$\begin{aligned} (15.35) \quad p_t(B|x) &= \mu(B; x)t + o(t), \quad x \notin B, \\ p_t(\{x\}|x) &= 1 - \mu(F; x)t + o(t). \end{aligned}$$

Now the problem is: Is there a unique Markov jump process fitting into (15.35)? Working backward from the results of the last section—we know that if there is a jump process satisfying (15.35), then it is appropriate to define

$$\lambda(x) = \mu(F; x), \quad p(B; x) = \mu(B; x)/\mu(F; x), \quad p(\{x\}; x) = 0,$$

and look for a process such that

$$P_x(T > t) = e^{-\lambda(x)t}, \quad P_x(X(T) \in B) = p(B; x).$$

Theorem 15.28 gives us the key to the construction of a pure jump process. Starting at  $x$ , we wait there a length of time  $T_0$  exponentially distributed with parameter  $\lambda(x)$ ; then independently of how long we wait, our first jump is to a position with distribution  $p(dy; x)$ . Now we wait at our new position  $y$ , time  $T_1$  independent of  $T_0$ , with distribution parameter  $\lambda(y)$ , etc. Note that these processes are very similar to the Poisson-like processes with independent increments. Heuristically, they are a sort of patched-together assembly of such processes, in the sense that at every point  $x$  the process behaves at that point like a Poisson-like process with parameter  $\lambda(x)$  and jump distribution given by  $p(B; x)$ .

At any rate, it is pretty clear how to proceed with the construction.

1) The space structure of this process is obtained by constructing a discrete Markov process  $X_0, X_1, X_2, \dots$ , moving under the transition probabilities  $p(B; x)$ , and starting from any point  $x$ .

2) *The time flow* of the process consists of slowing down or speeding up the rate at which the particle travels along the paths of the space structure.

For every  $n$ ,  $x \in F$ , construct random variables  $T_n(x)$  such that

- i)  $P(T_n(x) > t) = e^{-\lambda(x)t}$ ,
- ii)  $T_n(x)$  are jointly measurable in  $\omega, x$ ,
- iii) the processes  $(X_0, \dots)$ ,  $(T_0(x), x \in F)$ ,  $(T_1(x), x \in F), \dots$  are mutually independent.

The  $T_n(x)$  will serve as the waiting time in the state  $x$  after the  $n$ th jump. To see that joint measurability can be gotten, define on the probability space  $([0, 1], \mathcal{B}([0, 1]), dz)$  the variables  $T(z, \lambda) = -(1/\lambda) \log z$ . Thus  $P(T(z, \lambda) > t) = e^{-\lambda t}$ . Now define  $T_0(x) = T(z, \lambda(x))$  and take the cross-product space with the sample space for  $X_0, X_1, \dots$ . Similarly, for  $T_1(x), T_2(x), \dots$ .

For the process itself, proceed with

**Definition 15.36.** Define variables as follows:

$$\begin{aligned} R_0 &\equiv 0, & R_n &= T_0(X_0) + T_1(X_1) + \dots + T_{n-1}(X_{n-1}), \\ n^*(t) &= n, & \text{if } R_n &\leq t < R_{n+1}, \\ X(t) &= X_{n^*(t)}. \end{aligned}$$

In this definition  $R_n$  functions as the time of the  $n$ th jump.

**Theorem 15.37.** If  $n^*(t)$  is well-defined by 15.36 for all  $t$ , then  $X(t)$  is a pure jump Markov process with transition probabilities satisfying the given infinitesimal conditions.

*Proof.* This is a straightforward verification. The basic point is that given  $X(t) = x$ , and given, say, that we got to this space-time point in  $n$  steps, then the waiting time in  $x$  past  $t$  does not depend on how long has already been spent there; that is,

$$P(T_n(x) > \tau + \xi \mid T_n(x) > \xi) = e^{-\lambda(x)\tau}.$$

To show that the infinitesimal conditions are met, just show again that the probability of two jumps in time  $t$  is  $o(t)$ .

The condition that  $n^*(t)$  be well-defined is that

$$R_\infty = \sum_0^\infty T_k(X_k) = \infty \quad \text{a.s. } P_x, \quad \text{all } x.$$

This is a statement that at most a finite number of jumps can occur in every finite time interval. If  $P_x(R_\infty < \infty) > 0$ , there is no pure jump process that satisfies the infinitesimal conditions for all  $x \in F$ . However, even if  $R_\infty = \infty$  a.s., the question of uniqueness has been left open. Is there another Markov

process, not necessarily a jump process, satisfying the infinitesimal conditions? The answer, in regard to distribution, is No. The general result states that if  $P_x(R_\infty < \infty) = 0$ , all  $x \in F$ , then any Markov process  $\{X(t)\}$  satisfying the infinitesimal conditions (15.35) is a pure jump process and has the same distribution as the constructed process. For details of this, refer to Doob [39, pp. 266 ff.]. We content ourselves with the much easier assertion:

**Proposition 15.38.** *Any two pure jump processes having the same infinitesimal operator  $S$  have the same distribution.*

*Proof.* This is now almost obvious, because for both processes,  $T$  and  $X(T)$  have the same distribution. Therefore the sequence of variables  $X(T_0)$ ,  $X(T_0 + T_1)$ ,  $\dots$  has distribution governed by  $p(B; x)$ , and given this sequence, the jump times are sums of independent variables with the same distribution as the constructed variables  $\{R_n\}$ .

Let  $\{X(t)\}$  be the constructed process. Whether or not  $R_\infty = \infty$  a.s., define  $p_t^{(N)}(B | x)$  as the probability that  $X(t)$  reaches  $B$  in time  $t$  in  $N$  or fewer jumps. That is,

$$\begin{aligned} p_t^{(N)}(B | x) &= P_x(X(t) \in B, R_{N+1} > t) \\ &= \sum_{n=0}^N P_x(X_n \in B, R_n \leq t < R_{n+1}). \end{aligned}$$

For  $n \geq 1$ , and  $\tau < t$ ,

$$\begin{aligned} P_x(X_n \in B, T_0 + \dots + T_{n-1} \leq t < T_0 + \dots + T_n | X_1 = y, T_0 = \tau) \\ &= P_x(X_n \in B, T_1 + \dots + T_{n-1} \leq t - \tau < T_1 + \dots + T_n | X_1 = y) \\ &= P_y(X_{n-1} \in B, T_0 + \dots + T_{n-2} \leq t - \tau \\ &\quad < T_0 + \dots + T_{n-1}) \quad \text{a.s. } p(dy; x). \end{aligned}$$

The terms for  $n \geq 1$  vanish for  $\tau \geq t$ , and the zero term is

$$P_x(X_0 \in B, R_1 > t) = \chi_B(x)e^{-\lambda(x)t},$$

hence integrating out  $X_1, T_0$  gives

**Proposition 15.39**

$$p_t^{(N)}(B | x) = \chi_B(x)e^{-\lambda(x)t} + \int_0^t \int_{(x)} p_t^{(N-1)}(B | y) \mu(dy; x) e^{-\lambda(x)\tau} d\tau.$$

Letting  $N \rightarrow \infty$  gives another proof that for a pure jump process, integral equation (15.33) and hence the backwards equations are satisfied.

Define

$$(15.40) \quad \bar{p}_t(B | x) = \lim_N p_t^{(N)}(B | x).$$

The significance of  $\bar{p}_t(B | x) = \lim_N p_t^{(N)}(B | x)$  is that it is the probability of going from  $x$  to  $B$  in time  $t$  in a finite number of steps.

**Proposition 15.41.**  $\bar{p}_t(B | x)$  is the minimal solution of the backwards equations in the sense that if  $q_t(x)$  is any other solution satisfying

- 1)  $q_t(x) \geq 0$ ,
- 2)  $q_t(x) \rightarrow \chi_B(x)$  as  $t \rightarrow 0$ ,

then

$$\bar{p}_t(B | x) \leq q_t(x).$$

*Proof.* The backwards equation is

$$\frac{\partial}{\partial \tau} q_\tau(x) = -\lambda(x)q_\tau(x) + \int_{\{x\}^c} q_\tau(y)\mu(dy; x).$$

Multiply by  $e^{-\lambda(x)\tau}$ , integrate from 0 to  $t$ , and we recover the integral equation

$$q_t(x) = \chi_B(x)e^{-\lambda(x)t} + \int_0^t \int_{\{x\}^c} q_\tau(y)\mu(dy; x)e^{-\lambda(x)\tau} d\tau.$$

Assume  $q_t(x) \geq p_t^{(N)}(B | x)$ . Then substituting this inequality in the integral on the right,

$$\begin{aligned} q_t(x) &\geq \chi_B(x)e^{-\lambda(x)t} + \int_0^t \int_{\{x\}^c} p_\tau^{(N)}(B | y)\mu(dy; x)e^{-\lambda(x)\tau} d\tau \\ &= p_t^{(N+1)}(B | x). \end{aligned}$$

By the nonnegativity of  $q_t$ ,

$$q_t(x) \geq \chi_B(x)e^{-\lambda(x)t} = p_t^{(0)}(B | x).$$

Hence  $q_t(x) \geq \lim_N p_t^{(N)}(B | x) = \bar{p}_t(B | x)$ .

**Corollary 15.42.** If  $\{X(t)\}$  is a pure jump process, equivalently, if  $R_\infty = \infty$  a.s.  $P_x$ , all  $x \in F$ , then  $\{\bar{p}_t(B | x)\}$  are the unique set of transition probabilities satisfying the backwards equations.

## 7. EXPLOSIONS

If there are only a finite number of jumps in every finite time interval, then everything we want goes through—the forwards and backwards equations are satisfied and the solutions are unique. Therefore it becomes important to be able to recognize from the infinitesimal conditions when the resulting process will be pure jump. The thing that may foul the process up is unbounded  $\lambda(x)$ . The expected duration of stay in state  $x$  is given by  $E_x T = 1/\lambda(x)$ . Hence if  $\lambda(x) \rightarrow \infty$  anywhere, there is the possibility that the particle will move from state to state, staying in each one a shorter period of time. In the case where  $F$  represents the integers,  $\lambda(n)$  can go to  $\infty$  only if  $n \rightarrow \infty$ . In this case, we can have infinitely many jumps only if the particle

can move out to  $\infty$  in finite time. This is dramatically referred to as the possibility of *explosions* in the process. Perhaps the origin of this is in a population explosion model with pure birth,

$$\begin{aligned} p_t(j+1 | j) &= \lambda(j)t + o(t), \\ p_t(j | j) &= 1 - \lambda(j)t + o(t). \end{aligned}$$

Here the space structure is  $p(n+1; n) = 1$ ; the particle must move one step to the right each unit. Hence,  $X_n = n$  if  $X_0 = 0$ . Now

$$R_n = T_0(0) + \cdots + T_{n-1}(n-1)$$

is the time necessary to move  $n$  steps. And

$$E_0 R_\infty = \sum_0^\infty \frac{1}{\lambda(n)}.$$

If this sum is finite, then  $R_\infty < \infty$  a.s.  $P_0$ . This is also sufficient for  $P_j(R_\infty < \infty) = 1$ , for all  $j \in F$ . Under these circumstances the particle explodes out to infinity in finite time, and the theorems of the previous sections do not apply.

One criterion that is easy to derive is

**Proposition 15.43.** *A process satisfying the given infinitesimal conditions will be pure jump if and only if*

$$P_x \left( \sum_0^\infty \frac{1}{\lambda(X_n)} < \infty \right) = 0, \quad \text{all } x \in F.$$

*Proof.* For  $\sum_0^\infty T_n$  a sum of independent, exponentially distributed random variables with parameters  $\lambda_n$ ,  $\sum_0^\infty T_n < \infty$  a.s. iff  $\sum_0^\infty 1/\lambda_n < \infty$ . Because for  $s \geq 0$ ,

$$E e^{-s(T_1 + \cdots + T_n)} = \frac{1}{\prod_1^n (1 + s/\lambda_k)}.$$

Verify that the infinite product on the right converges to a finite limit iff  $\sum_0^\infty 1/\lambda_n < \infty$ , and apply 8.36 and the continuity theorem given in Section 13, Chapter 8. Now note that given  $(X_0, X_1, \dots)$ ,  $R_\infty$  is a sum of such variables with parameters  $\lambda(X_n)$ .

**Corollary 15.44.** *If  $\sup_{x \in F} \lambda(x) < \infty$ , then  $\{X(t)\}$  is pure jump.*

Note that for a pure birth process  $\sum_0^\infty 1/\lambda(n) = \infty$  is both necessary and sufficient for the process to be pure jump. For  $F$  the integers, another obvious sufficient condition is that every state be recurrent under the space structure. Conversely, consider

$$E_j \left( \sum_0^\infty \frac{1}{\lambda(X_n)} \right).$$

Let  $N(k)$  be the number of entries that the sequence  $X_0, X_1, \dots$  makes into state  $k$ . Then

$$E_j \left( \sum_0^{\infty} \frac{1}{\lambda(X_n)} \right) = \sum_k \frac{E_j N(k)}{\lambda(k)}.$$

If this is finite, then certainly there will be an explosion. Time-continuous birth and death processes are defined as processes on the integers such that each jump can be only to adjacent states, so that

$$\begin{aligned} p_t(j+1 | j) &= b(j)t + o(t), \\ p_t(j-1 | j) &= d(j)t + o(t), \\ p_t(j | j) &= 1 - (b(j) + d(j))t + o(t). \end{aligned}$$

For a birth and death process with no absorbing states moving on the non-negative integers,  $E_0(N(k)) = 1 + M(k)$ , where  $M(k)$  is the expected number of returns to  $k$  given  $X_0 = k$ . Then, as above,

$$E_0 \left( \sum_0^{\infty} \frac{1}{\lambda(X_n)} \right) = \sum_k \frac{1 + M(k)}{\lambda(k)}.$$

The condition that this latter sum be infinite is both necessary and sufficient for no explosions [12].

Another method for treating birth and death processes was informally suggested to us by Charles Stone. Let  $F$  be the nonnegative integers with no absorbing states. Let  $\tau_n^*$  be the first passage time from state 1 to state  $n$ , and  $\tau_n^*$  the first passage time from state  $n$  to state  $n+1$ . The  $\tau_1^*, \tau_2^*, \dots$  are independent,  $\tau_n^* = \sum_1^{n-1} \tau_k^*$ .

**Proposition 15.45.**  $\tau_{\infty}^* = \lim_n \tau_n^*$  is finite a.s. if and only if  $\sum_1^{\infty} E\tau_k^* < \infty$ .

*Proof.* Let  $T_k$  be the duration of first stay in state  $k$ , then  $\tau_k^* \geq T_k$ . Further  $\sum_1^{\infty} T_k < \infty$  a.s. iff  $\sum_1^{\infty} ET_k < \infty$  or  $\sum_1^{\infty} 1/\lambda(k) < \infty$ . Hence if  $\inf_k \lambda(k) = 0$ , both  $\sum_1^{\infty} \tau_k^*$  and  $\sum_1^{\infty} E\tau_k^*$  are infinite. Now assume  $\inf_k \lambda(k) = \delta > 0$ . Given any succession of states  $X_0 = n, X_1, \dots, X_m = n+1$  leading from  $n$  to  $n+1$ ,  $\tau_k^*$  is a sum of independent, exponentially distributed random variables  $T_0 + \dots + T_m$ , and  $\sigma^2(T_0 + \dots + T_m) = \sigma^2(T_0) + \dots + \sigma^2(T_m) \leq \delta^{-1}E(T_0 + \dots + T_m)$ . Hence

$$\sigma^2(\tau_k^*) \leq \delta^{-1}E\tau_k^*.$$

If  $\sum_1^{\infty} \tau_k^*$  converges a.s., but  $\sum_1^{\infty} E\tau_k^* = \infty$ , then for  $0 < \epsilon < 1$ ,

$$P \left( \left| \sum_1^n (\tau_k^* - E\tau_k^*) \right| > \epsilon \sum_1^n E\tau_k^* \right) \rightarrow 1.$$

Applying Chebyshev's inequality to this probability gives a contradiction which proves the proposition.

### Problems

6. Show that  $\alpha(k) = E\tau_k^*$  satisfies the difference equation

$$\alpha(k) = p(k) \cdot \frac{1}{\lambda(k)} + q(k) \left[ \frac{1}{\lambda(k)} + \alpha(k-1) + \alpha(k) \right], \quad k \geq 1,$$

where

$$p(k) = b(k)/\lambda(k), \quad q(k) = d(k)/\lambda(k),$$

or

$$p(k)\alpha(k) = \frac{1}{\lambda(k)} + q(k)\alpha(k-1), \quad k \geq 1.$$

Deduce conditions on  $p(k)$ ,  $q(k)$ ,  $\lambda(k)$  such that there are no explosions. (See [87].)

7. Discuss completely the explosive properties of a birth and death process with  $\{0\}$  a reflecting state and

$$b(n) = b\varphi(n), \quad d(n) = d\varphi(n).$$

### 8. NONUNIQUENESS AND BOUNDARY CONDITIONS

If explosions are possible, then  $\bar{p}_t(F|x) < 1$  for some  $x$ ,  $t$ , and the process is not uniquely determined by the given infinitesimal conditions. The nature of the nonuniqueness is that the particle can *reach* points on some undefined "boundary" of  $F$  not included in  $F$ . Then to completely describe the process it is necessary to specify its evolution from these boundary points. This is seen most graphically when  $F$  is the integers. If  $P_j(R_\infty < \infty) > 0$  for some  $j$ , then we have to specify what the particle will do once it reaches  $\infty$ . One possible procedure is to add to  $F$  a state denoted  $\{\infty\}$  and to specify transition probabilities from  $\{\infty\}$  to  $j \in F$ . For example, we could make  $\{\infty\}$  an absorbing state, that is,  $p_t(\{\infty\} | \{\infty\}) = 1$ . An even more interesting construction consists of specifying that once the particle reaches  $\{\infty\}$  it instantaneously moves into state  $k$  with probability  $Q(k)$ . This is more interesting in that it is not necessary to adjoin an extra state  $\{\infty\}$  to  $F$ .

To carry out this construction, following Chung [16], define

$$p_t^{(0)}(j|k) = \bar{p}_t(j|k) = P_k(X(t) = j, R_\infty \geq t).$$

Now look at the probability  $p_t^{(1)}(j|k)$  that  $k \rightarrow j$  in time  $t$  with exactly one passage to  $\{\infty\}$ . To compute this, suppose that  $R_\infty = \tau$ ; then the particle moves immediately to state  $l$  with probability  $Q(l)$ , and then must go from  $l$  to  $j$  in time  $t - \tau$  with no further excursions to  $\{\infty\}$ . Hence, denoting  $H_k(d\tau) = P_k(R_\infty \in d\tau)$ ,

$$p_t^{(1)}(j|k) = \int_0^t \left[ \sum_l p_{t-\tau}^{(0)}(j|l) Q(l) \right] H_k(d\tau).$$



Similarly, the probability  $p^{(n)}(j | k)$ , of  $k \rightarrow j$  in time  $t$  with exactly  $n$  passages to  $\{\infty\}$  is given by

$$p_t^{(n)}(j | k) = \int_0^t \left[ \sum_l p_{t-\tau}^{(n-1)}(j | l) Q(l) \right] H_k(d\tau).$$

Now define

$$p_t(j | k) = \sum_{n=0}^{\infty} p_t^{(n)}(j | k).$$

**Proposition 15.46.**  $p_t(j | k)$  as defined above satisfies

- 1)  $\sum_j p_t(j | k) = 1$ ,
- 2) the Chapman-Kolmogorov equations, and
- 3) the backwards equations.

*Proof.* Left to reader.

**Remark.**  $p_t(j | k)$  does not satisfy the forwards equations. See Chung [16, pp. 224 ff.].

The process constructed the above way has the property that

$$p_t(j | k) = p_t^{(0)}(j | k) + o(t).$$

This follows from noting that

$$p_t(j | k) = p_t^{(0)}(j | k) + \int_0^t \left( \sum_l p_{t-\tau}(j | l) Q(l) \right) H_k(d\tau).$$

The integral term is dominated by  $P_k(R_\infty < t)$ . This is certainly less than the probability of two jumps in time  $t$ , hence is  $o(t)$ . Therefore, no matter what  $Q(l)$  is, all these processes have the specified infinitesimal behavior. This leads to the observation (which will become more significant in the next chapter), that *if it is possible to reach a "boundary" point, then boundary conditions must be added to the infinitesimal conditions in order to specify the process.*

## 9. RESOLVENT AND UNIQUENESS

Although  $S$  with domain  $\mathcal{D}(S)$  does not determine the process uniquely, this can be fixed up with a more careful and restrictive definition of the domain of  $S$ . In this section the processes dealt with will be assumed to have standard stationary transition probabilities, but no restrictions are put on their sample paths.

**Definition 15.47.** Say that functions  $\varphi_t(x)$  converge boundedly pointwise to  $\varphi(x)$  on some subset  $A$  of their domain as  $t \rightarrow 0$  if

$$\text{i) } \lim_{t \rightarrow 0} \varphi_t(x) = \varphi(x), \quad \text{all } x \in A,$$

ii)  $\sup_{x \in A} |\varphi_t(x)| \leq M < \infty$ , for all  $t$  sufficiently small.

Denote this by  $\varphi_t(x) \xrightarrow{bp} \varphi(x)$  on  $A$ .

Let  $\mathcal{L}$  be any class of bounded  $\mathcal{B}_1(F)$ -measurable functions. Then we use

**Definition 15.48.**  $\mathcal{D}(S, \mathcal{L})$  consists of all functions  $f(x)$  in  $\mathcal{L}$  such that

$$\frac{E_x(f(X(t))) - f(x)}{t}$$

converges boundedly pointwise on  $F$  to a function in  $\mathcal{L}$ .

We plan to show that with an appropriate choice of  $\mathcal{L}$ , that corresponding to a given  $S$ ,  $\mathcal{D}(S, \mathcal{L})$ , there is at most one process. In the course of this, we will want to integrate functions of the type  $E_x f(X(t))$ , so we need

**Proposition 15.49.** For  $f(x)$  bounded and measurable on  $F$ ,  $\varphi(x, t) = E_x f(X(t))$  is jointly measurable in  $(x, t)$ , with respect to  $\mathcal{B}_1(F) \times \mathcal{B}_1([0, \infty))$ .

*Proof.* Take  $f(x)$  bounded and continuous. Since  $\{X(t)\}$  is continuous in probability from the right, the function  $\varphi(x, t) = E_x f(X(t))$  is continuous in  $t$  from the right and  $\mathcal{B}_1(F)$ -measurable in  $x$  for  $t$  fixed. Consider the approximation

$$\varphi_n(x, t) = \sum_{k=0}^{\infty} \chi_{[k/2^n, (k+1)/2^n)}(t) \varphi(x, (k+1)/2^n).$$

By the right-continuity,  $\varphi_n(x, t) \rightarrow \varphi(x, t)$ . But  $\varphi_n(x, t)$  is jointly measurable, therefore so is  $\varphi(x, t)$ . Now consider the class of  $\mathcal{B}_1(F)$ -measurable functions  $f(x)$  such that  $|f(x)| \leq 1$  and the corresponding  $\varphi(x, t)$  is jointly measurable. This class is closed under pointwise convergence, and contains all continuous functions bounded by one. Hence it contains all bounded measurable functions bounded by one.

**Definition 15.50.** The resolvent is defined as

$$R_\lambda(B | x) = \int_0^\infty e^{-\lambda t} p_t(B | x) dt, \quad \lambda > 0$$

for any  $B \in \mathcal{B}_1(F)$ ,  $x \in F$ .

It is easy to check that  $R_\lambda(B | x)$  is a bounded measure on  $\mathcal{B}_1(F)$  for fixed  $x$ . Furthermore, by 15.49 and the Fubini theorem,  $R_\lambda(B | x)$  is  $\mathcal{B}_1(F)$ -measurable in  $x$  for  $B \in \mathcal{B}_1(F)$  fixed. Denote, for  $f$  bounded and  $\mathcal{B}_1(F)$ -measurable,

$$(R_\lambda f)(x) = \int f(y) R_\lambda(dy | x) \quad \text{and} \quad (T_t f)(x) = \int f(y) p_t(dy | x).$$

Then, using the Fubini theorem to justify the interchange,

$$R_\lambda f = \int_0^\infty e^{-\lambda t} (T_t f) dt.$$

Take  $\mathfrak{G}$  to be the set of all bounded  $\mathfrak{B}_1(F)$ -measurable functions  $f(x)$  such that  $(T_t f)(x) \rightarrow f(x)$  as  $t \rightarrow 0$  for every  $x \in F$ . Note that  $\mathfrak{D}(S, \mathfrak{G}) \subset \mathfrak{G}$ .

**Theorem 15.51.** *If  $f \in \mathfrak{G}$ , then  $R_\lambda f$  is in  $\mathfrak{D}(S, \mathfrak{G})$  and*

$$(1) \quad (\lambda - S)(R_\lambda f) = f.$$

*If  $f$  is in  $\mathfrak{D}(S, \mathfrak{G})$ , then*

$$(2) \quad R_\lambda((\lambda - S)f) = f.$$

*Proof.* If  $f$  is in  $\mathfrak{G}$ , then since  $T_t$  and  $R_\lambda$  commute, the bounded convergence theorem can be applied to  $T_t(R_\lambda f) = R_\lambda(T_t f)$  to establish  $R_\lambda f \in \mathfrak{G}$ . Write

$$T_t(R_\lambda f) = T_t \int_0^\infty e^{-\lambda \tau} (T_\tau f) d\tau = \int_0^\infty e^{-\lambda \tau} (T_t(T_\tau f)) d\tau.$$

The Chapman-Kolmogorov equations imply  $T_t(T_\tau f) = T_{t+\tau} f$ , so

$$T_t(R_\lambda f) = e^{\lambda t} \left[ R_\lambda f - \int_0^t e^{-\lambda \tau} (T_\tau f) d\tau \right].$$

From this, denoting

$$S_t f = \frac{T_t f - f}{t},$$

we get

$$(15.52) \quad S_t(R_\lambda f) = \left( \frac{e^{-\lambda t} - 1}{t} \right) R_\lambda f - \frac{e^{\lambda t}}{t} \int_0^t e^{-\lambda \tau} (T_\tau f) d\tau.$$

Using  $\|f\| = \sup |f(x)|$ ,  $x \in F$ , we have

$$\begin{aligned} \|S_t(R_\lambda f)\| &\leq \frac{1 - e^{-\lambda t}}{\lambda t} \|f\| + \frac{e^{\lambda t}(1 - e^{-\lambda t})}{\lambda t} \|f\|; \\ &\leq c \|f\|, \quad 0 < t \leq 1. \end{aligned}$$

As  $t$  goes to zero,  $(e^{-\lambda t} - 1)/t \rightarrow -\lambda$ . As  $\tau$  goes to zero,  $(T_\tau f)(x) \rightarrow f(x)$ . Using these in (15.52) completes the proof of the first assertion.

Now take  $f$  in  $\mathfrak{D}(S, \mathfrak{G})$ ; by the bounded convergence theorem,

$$R_\lambda((\lambda - S)f) = \lim_{t \rightarrow 0} R_\lambda((\lambda - S_t)f).$$

Note that  $R_\lambda$  and  $S_t$  commute, so

$$R_\lambda((\lambda - S)f) = \lim_{t \rightarrow 0} (\lambda - S_t)(R_\lambda f).$$

By part (1) of this theorem,  $R_\lambda f$  is in  $\mathcal{D}(S, \mathcal{G})$ , hence

$$\lim_{t \rightarrow 0} (\lambda - S_t)(R_\lambda f) = (\lambda - S)(R_\lambda f) = f.$$

The purpose of this preparation is to prove

**Theorem 15.53.** *There is at most one set of standard transition probabilities corresponding to given  $S, \mathcal{D}(S, \mathcal{G})$ .*

*Proof.* Suppose there are two different sets,  $p_t^{(1)}$  and  $p_t^{(2)}$  leading to resolvents  $R_\lambda^{(1)}$  and  $R_\lambda^{(2)}$ . For  $f$  in  $\mathcal{G}$ , let

$$g = (R_\lambda^{(1)} - R_\lambda^{(2)})f.$$

Then, by 15.51(1),

$$(\lambda - S)g = 0.$$

But  $g \in \mathcal{D}(S, \mathcal{G})$ , so use 15.51(2) to get

$$R_\lambda^{(1)}((\lambda - S)g) = g.$$

Therefore  $g$  is zero. Thus for all  $f \in \mathcal{G}$ ,  $R_\lambda^{(1)}f = R_\lambda^{(2)}f$ . Since  $\mathcal{G}$  includes all bounded continuous functions, for any such function, and for all  $\lambda > 0$ ,

$$\int_0^\infty e^{-\lambda t} (T_t^{(1)}f) dt = \int_0^\infty e^{-\lambda t} (T_t^{(2)}f) dt.$$

By the uniqueness theorem for Laplace transforms (see [140], for example)  $(T_t^{(1)}f)(x) = (T_t^{(2)}f)(x)$  almost everywhere ( $dt$ ). But both these functions are continuous from the right, hence are identically equal. Since bounded continuous functions separate,  $p_t^{(1)}(B | x) \equiv p_t^{(2)}(B | x)$ .

The difficulty with this result is in the determination of  $\mathcal{D}(S, \mathcal{G})$ . This is usually such a complicated procedure that the uniqueness theorem 15.53 above has really only theoretical value. Some examples follow in these problems.

### Problems

8. For the transition probabilities constructed and referred to in 15.46, show that a necessary condition for  $f(j)$  to be in  $\mathcal{D}(S, \mathcal{G})$  is

$$\sum_j f(j)Q(j) = 0.$$

9. Show that for any  $\lambda', \lambda'' > 0$ ,

$$R_{\lambda'} f = R_{\lambda'} f + (\lambda' - \lambda'')R_{\lambda'}(R_{\lambda''} f).$$

Use this to show that the set  $\mathcal{R}_\lambda$  consisting of all functions  $\{R_\lambda f\}$ ,  $f$  in  $\mathcal{G}$ , does not depend on  $\lambda$ . Use 15.51 to show that  $\mathcal{R}_\lambda = \mathcal{D}(S, \mathcal{G})$ .

10. For Brownian motion, show that the resolvent has a density  $r_\lambda(y|x)$  with respect to  $dy$  given by

$$r_\lambda(y|x) = (1/\sqrt{2\lambda})e^{-\sqrt{2\lambda}|y-x|}.$$

11. Let  $\mathcal{C}$  be the class of all bounded continuous functions on  $R^{(1)}$ . Use the identity in Problem 9, and the method of that problem to show that  $\mathcal{D}(S, \mathcal{C})$  for Brownian motion consists of all functions  $f$  in  $\mathcal{C}$  such that  $f''(x)$  is in  $\mathcal{C}$ .

12. For a pure jump process, show that if  $\sup \lambda(x) < \infty$ , then  $\mathcal{D}(S, \mathcal{G})$  consists of all bounded  $\mathcal{B}_1(F)$ -measurable functions.

## 10. ASYMPTOTIC STATIONARITY

Questions concerning the asymptotic stationarity of a Markov process  $\{X(t)\}$  can be formulated in the same way as for discrete time chains. In particular,

**Definition 15.54.**  $\bar{\pi}(dx)$  on  $\mathcal{B}_1(F)$  will be called a *stationary initial distribution* for the process if for every  $B \in \mathcal{B}_1(F)$  and  $t > 0$ ,

$$\bar{\pi}(B) = \int p_t(B|x)\bar{\pi}(dx).$$

Now ask, when do the probabilities  $p_t(B|x)$  converge as  $t \rightarrow \infty$  to some stationary distribution  $\bar{\pi}(B)$  for all  $x \in F$ ? Interestingly enough, the situation here is less complicated than in discrete time because there is no periodic behavior. We illustrate this for  $X(t)$  a pure jump process moving on the integers. Define the times between successive returns to state  $k$  by

$$\tau_1^* = \inf \{t; X(t) = k, \exists \tau < t \text{ such that } X(\tau) \neq k\},$$

$$\tau_2^* = \inf \{t; X(t + \tau_1^*) = k, \exists \tau_1^* < \tau < t + \tau_1^* \text{ such that } X(\tau) \neq k\},$$

and so forth. By the strong Markov property,

**Proposition 15.55.** If  $P_k(\tau_1^* < \infty) = 1$ , then the  $\tau_1^*, \tau_2^*, \dots$  are independent and identically distributed.

If  $P_k(\tau_1^* < \infty) < 1$ , then the state  $k$  is called *transient*.

To analyze the asymptotic behavior of the transition probabilities, use

$$\{X(t) = k\} = \bigcup_{n=0}^{\infty} \{\tau_1^* + \cdots + \tau_n^* \leq t, \tau_1^* + \cdots + \tau_n^* + T_n(k) > t\},$$

where  $T_n(k)$  is the duration of stay in state  $k$  after the  $n$ th return. It is independent of  $\tau_1^*, \dots, \tau_n^*$ , and  $P_k(T_n(k) > t) = e^{-\lambda(k)t}$ . So

$$p_t(k|k) = e^{-\lambda(k)t} + \sum_{n=1}^{\infty} P_k(\tau_1^* + \cdots + \tau_n^* \leq t, \tau_1^* + \cdots + \tau_n^* + T_n(k) > t).$$

Put

$$R(t) = \sum_1^{\infty} P_k(\tau_1^* + \cdots + \tau_n^* \leq t);$$

then

$$p_t(k | k) = e^{-\lambda(k)t} + \int_0^t P_k(T_0(k) > \xi) R(t - d\xi).$$

Argue that  $\tau_1^*$  is nonlattice because

$$\tau_1^* = T_0(k) + \tau_1^*,$$

where  $\tau_1^*$  is the time from the first exit to the first return. By the strong Markov property,  $T_0(k)$ ,  $\tau_1^*$  are independent. Finally, note that  $T_0(k)$  has a distribution absolutely continuous with respect to Lebesgue measure; hence so does  $\tau_1^*$ .

Now apply the renewal theorem 10.8. As  $t \rightarrow \infty$ ,

$$R(t - d\xi) \xrightarrow{w} \frac{d\xi}{E\tau_1^*}.$$

Conclude that

$$\lim_t p_t(k | k) = \frac{1}{E\tau_1^*} \int_0^{\infty} e^{-\lambda(k)t} dt.$$

Hence

**Proposition 15.56.** *Let  $T$  be the first exit time from state  $k$ ,  $\tau_1^*$  the time of first return. If  $E\tau_1^* < \infty$ , then*

$$(15.57) \quad p_t(k | k) \rightarrow \frac{ET}{E\tau_1^*} = \pi(k).$$

The following problems concern the rest of the problem of asymptotic convergence.

### Problems

**13.** Let all the states communicate under the space structure given by  $p(j; k)$ , and let the expected time of first return be finite for every state. Show that

1) For any  $k, j$ ,

$$\lim_t p_t(j | k) = \pi(j),$$

where  $\pi(j)$  is defined by (15.57).

2) If  $\tilde{\pi}(k)$  is the stationary initial distribution under  $p(j; k)$ , that is,

$$\tilde{\pi}(j) = \sum_k p(j; k) \tilde{\pi}(k),$$

then

$$\pi(j) = \alpha \tilde{\pi}(j) / \lambda(j),$$

where  $\alpha$  is normalizing constant.

14. Show that if  $k$  is a transient state, then  $p_t(k | k)$  goes to zero exponentially fast.
15. Show that  $\pi$  is a stationary initial distribution for a process having standard stationary transition probabilities if and only if

$$\int g(y)\pi(dy) = 0$$

for all  $g$  such that there exists an  $f$  in  $\mathcal{D}(S, \mathcal{G})$  with  $g = Sf$ .

## NOTES

K. L. Chung's book [16] is an excellent reference for the general structure of time-continuous Markov chains with a countable state space. Even with this simple a state space, the diversity of sample path behavior of processes with standard stationary transition probabilities is dazzling. For more general state spaces and for jump processes in particular, see Doob's book [39, Chap. 6]. For a thoroughly modern point of view, including discussion of the strong Markov property and the properties of  $S$ ,  $\mathcal{D}(S, \mathcal{G})$  and the resolvent, see Dynkin [44, especially Vol. I].

The fundamental work in this field started with Kolmogorov [93, 1931]. The problems concerning jump processes were treated analytically by Pospíšil [117, 1935–1936] and Feller in 1936, but see Feller [52] for a fuller treatment. Doebelin [26, 1939] had an approach closer in spirit to ours. Doob [33 in 1942] carried on a more extended study of the sample path properties. The usefulness of the resolvent and the systematic study of the domain of  $S$  were introduced by Feller [57, 1952]. His idea was that the operators  $\{T_t\}$ ,  $t \geq 0$ , formed a semi-group, hence methods for analyzing semi-groups of operators could be applied to get useful results.

There is an enormous literature on applications of pure jump Markov processes, especially for those with a countable state space. For a look at some of those, check the books by Bharucha-Reid [3], T. E. Harris [67], N. T. J. Bailey [2], and T. L. Saaty [120]. An extensive reference to both theoretical and applied sources is the Bharucha-Reid book.

## CHAPTER 16

## DIFFUSIONS

### 1. THE ORNSTEIN-UHLENBECK PROCESS

In Chapter 12, the Brownian motion process was constructed as a model for a microscopic particle in liquid suspension. We found the outstanding nonreality of the model was the assumption that increments in displacement were independent—ignoring the effects of the velocity of the particle at the beginning of the incremental time period. We can do better in the following way:

Let  $V(t)$  be the *velocity* of a particle of mass  $m$  suspended in liquid. Let  $\Delta V = V(t + \Delta t) - V(t)$ , so that  $m \Delta V$  is the change in momentum of the particle during time  $\Delta t$ . The basic equation is

$$(16.1) \quad m \Delta V = -\beta V \Delta t + \Delta M.$$

Here  $-\beta V$  is the viscous resistance force, so  $-\beta V \Delta t$  is the loss in momentum due to viscous forces during  $\Delta t$ .  $\Delta M$  is the momentum transfer due to molecular bombardment of the particle during time  $\Delta t$ .

Let  $M(t)$  be the momentum transfer up to time  $t$ . Normalize arbitrarily to  $M(0) = 0$ . Assume that

- i)  $M(t + \Delta t) - M(t)$  is independent of  $\mathcal{F}(M(\tau), \tau \leq t)$ ,
- ii) the distribution of  $\Delta M$  depends only on  $\Delta t$ ,
- iii)  $M(t)$  is continuous in  $t$ .

The third assumption may be questionable if one uses a hard billiard-ball model of molecules. But even in this case we reason that the jumps of  $M(t)$  would have to be quite small unless we allowed the molecules to have enormous velocities. At any rate (iii) is not unreasonable as an approximation.

But (i), (ii), (iii) together characterize  $M(t)$  as a Brownian motion. The presence of drift in  $M(t)$  would put a  $\mu \Delta t$  term on the right-hand side of (16.1). Such a term corresponds to a constant force field, and would be useful, for example, in accounting for a gravity field. However, we will assume no constant force field exists, and set  $EM(t) = 0$ . Put  $EM^2(t) = \sigma^2 t$ ; hence  $M(t) = \sigma X(t)$ , where  $X(t)$  is normalized Brownian motion. Equation (16.1) becomes

$$(16.2) \quad m \Delta V = -\beta V \Delta t + \sigma \Delta X.$$



Doing what comes naturally, we divide by  $\Delta t$ , let  $\Delta t \rightarrow 0$  and produce the Langevin equation

$$(16.3) \quad m \frac{dV}{dt} = -\beta V + \sigma \frac{dX}{dt}.$$

The difficulty here is amusing: We know from Chapter 12 that  $dX/dt$  exists nowhere. So (16.3) makes no sense in any orthodox way. But look at this: Write it as

$$\frac{d}{dt}(e^{\alpha t}V(t)) = \gamma e^{\alpha t} \frac{dX(t)}{dt},$$

where  $\alpha = \beta/m$ ,  $\gamma = \sigma/m$ . Assume  $V(0) = 0$  and integrate from 0 to  $t$  to get

$$e^{\alpha t}V(t) = \gamma \int_0^t e^{\alpha \tau} dX(\tau).$$

Do an integration by parts on the integral,

$$e^{\alpha t}V(t) = \gamma e^{\alpha t}X(t) - \gamma \alpha \int_0^t X(\tau) e^{\alpha \tau} d\tau.$$

Now the integral appearing is for each  $\omega$  just the integral of a continuous function and makes sense. Thus the expression for  $V(t)$  given by

$$V(t) = \gamma \int_0^t e^{\alpha(t-\tau)} dX(\tau)$$

can be well defined by this procedure, and results in a process with continuous sample paths.

To get a more appealing derivation, go back to (16.2). Write it as

$$\Delta(e^{\alpha t}V) = \gamma e^{\alpha t} \Delta X + \delta(\Delta t),$$

where  $\delta(\Delta t) = o(\Delta t)$  because by (16.2),  $V(t)$  is continuous and bounded in every finite interval. By summing up, write this as

$$e^{\alpha t}V(t) \simeq \gamma \sum_{k=0}^{n-1} e^{\alpha t_k^{(n)}} (X(t_{k+1}^{(n)}) - X(t_k^{(n)})),$$

where  $0 = t_0^{(n)} < \cdots < t_n^{(n)} = t$  is a partition  $\mathcal{P}_n$  of  $[0, t]$ . If the limit of the right-hand side exists in some decent way as  $\|\mathcal{P}_n\| \rightarrow 0$ , then it would be very reasonable to define  $e^{\alpha t}V(t)$  as this limit. Replace the integration by parts in the integral by a similar device for the sum,

$$\sum_{k=0}^{n-1} e^{\alpha t_k} (X(t_{k+1}) - X(t_k)) = e^{\alpha t} X(t) - \sum_{k=0}^{n-1} (e^{\alpha t_{k+1}} - e^{\alpha t_k}) X(t_{k+1}).$$

The second sums are the Riemann-Stieltjes sums for the integral  $\int_0^t X(\tau) d(e^{\alpha \tau})$ .

For every sample path, they converge to the integral. Therefore:

**Definition 16.4.** The Ornstein-Uhlenbeck process  $V(t)$  normalized to be zero at  $t = 0$ , is defined as

$$V(t) = \gamma \int_0^t e^{-\alpha(t-\tau)} dX(\tau),$$

where the integral is the limit of the approximating sums for every path.

**Proposition 16.5.**  $V(t)$  is a Gaussian process with covariance

$$EV(s)V(t) = \rho(e^{-\alpha|s-t|} - e^{-\alpha(s+t)}),$$

where  $\rho = \gamma^2/2\alpha$ .

*Proof.* That  $V(t)$  is Gaussian follows from its being the limit of sums  $\sum_k \varphi(t_k) \Delta_k X$ , where the  $\Delta_k X = X(t_{k+1}) - X(t_k)$  are independent, normally-distributed random variables. To get  $\Gamma(s, t)$ , take  $s > t$ , put

$$0 = t_0 < \cdots < t_n = t < t_{n+1} < \cdots < t_m = s.$$

Write

$$V(s) \simeq \gamma \sum_0^m e^{-\alpha(s-t_k)} \Delta_k X, \quad V(t) \simeq \gamma \sum_0^n e^{-\alpha(t-t_k)} \Delta_k X.$$

Use  $E(\Delta_k X)(\Delta_j X) = 0$  if  $k \neq j$ ,  $E(\Delta_k X)^2 = t_{k+1} - t_k$ , to get

$$EV(s)V(t) \simeq \gamma^2 \sum_0^n e^{-\alpha(s+t)+2\alpha t_k} (t_{k+1} - t_k).$$

Going to the limit, we get

$$EV(s)V(t) = \gamma^2 e^{-\alpha(s+t)} \int_0^t e^{2\alpha\tau} d\tau.$$

As  $t \rightarrow \infty$ ,  $EV(t)^2 \rightarrow \rho$ , so  $V(t) \xrightarrow{\mathcal{D}} V(\infty)$ , where  $V(\infty)$  is  $\mathcal{N}(0, \rho)$ . What if we start the process with this limiting distribution? This would mean that the integration of the Langevin equation would result in

$$e^{\alpha t} V_1(t) - V_1(0) = \gamma \int_0^t e^{-\alpha(t-\tau)} dX(\tau).$$

Define the stationary Ornstein-Uhlenbeck process by

$$(16.6) \quad Y(t) = V(t) + e^{-\alpha t} V_1(0),$$

where  $V_1(0)$  is  $\mathcal{N}(0, \rho)$  and independent of  $\mathcal{F}(V(t), t \geq 0)$ .

**Proposition 16.7.**  $Y(t)$  is a stationary Gaussian process with covariance  $\Gamma(s, t) = \rho e^{-\alpha|s-t|}$ .

*Proof.* Direct computation.

**Remark.** Stationarity has not been defined for continuous parameter processes, but the obvious definition is that all finite-dimensional distributions remain invariant under a time shift. For *Gaussian processes* with zero means, stationarity is equivalent to  $\Gamma(s, t) = \varphi(|s - t|)$ .

The additional important properties of the Ornstein-Uhlenbeck process are:

**Proposition 16.8.**  $Y(t)$  is a Markov process with stationary transition probabilities having all sample paths continuous.

*Proof.* Most of 16.8 follows from the fact that, for  $\tau \geq 0$ ,  $Y(t + \tau) - e^{-\alpha t}Y(\tau)$  is independent of  $\mathcal{F}(Y(s), s \leq \tau)$ . To prove this, it is necessary only to check the covariance

$$\begin{aligned} E[(Y(t + \tau) - e^{-\alpha t}Y(\tau))Y(s)] &= \Gamma(t + \tau, s) - e^{-\alpha t}\Gamma(\tau, s) \\ &= 0, \quad s \leq \tau. \end{aligned}$$

Now,

$$\begin{aligned} P(Y(t + \tau) \in A \mid Y(\tau) = x, Y(s), s \leq \tau) \\ &= P(Y(t + \tau) - e^{-\alpha t}Y(\tau) \in A - e^{-\alpha t}x \mid Y(\tau) = x, Y(s), s \leq \tau) \\ &= P(Y(t + \tau) - e^{-\alpha t}Y(\tau) \in A - e^{-\alpha t}x). \end{aligned}$$

The random variable  $Y(t + \tau) - e^{-\alpha t}Y(\tau)$  is normal with mean zero, and

$$\begin{aligned} E(Y(t + \tau) - e^{-\alpha t}Y(\tau))^2 \\ &= E[(Y(t + \tau) - e^{-\alpha t}Y(\tau))(Y(t + \tau))] = \rho(1 - e^{-2\alpha t}). \end{aligned}$$

Thus  $p_t(\cdot \mid x)$  has the distribution of

$$\mathcal{N}(e^{-\alpha t}x, \rho(1 - e^{-2\alpha t})).$$

The continuity of paths follows from the definition of  $V(t)$  in terms of an integral of  $X(t)$ .

## Problems

1. Show that if a process is Gaussian, stationary, Markov, and continuous in probability, then it is of the form  $Y(t) + c$ , where  $Y(t)$  is an Ornstein-Uhlenbeck process.

2. Let  $Z$  be a vector-valued random variable taking values in  $R^{(m)}$ ,  $m \geq 2$ . Suppose that the components of  $Z$ ,  $(Z_1, Z_2, \dots, Z_m)$ , are independent and identically distributed with a symmetric distribution. Suppose also that the components have the same property under all other orthogonal coordinate systems gotten from the original one by rotation. Show that  $Z_1, \dots, Z_m$  are  $\mathcal{N}(0, \sigma^2)$ .

**Remark.** The notable result of this problem is that any model for Brownian motion in three dimensions leads to variables normally distributed providing the components of displacement of velocity along the different axes are independent and identically distributed (symmetry is not essential, see Kac [79]) irrespective of which orthogonal coordinate system is selected. However, it does not follow from this that the process must be Gaussian.

## 2. PROCESSES THAT ARE LOCALLY BROWNIAN

In the spirit of the Langevin approach of the last section, if  $Y(t)$  is Brownian motion with drift  $\mu$ , variance  $\sigma^2$ , then write

$$\Delta Y = \mu \Delta t + \sigma \Delta X.$$

The same integration procedure as before would then result in a process  $Y(t)$  which would be, in fact, Brownian motion with parameters  $\mu, \sigma$ . To try to get more general Markov processes with continuous paths, write

$$(16.9) \quad \Delta Y = \mu(Y) \Delta t + \sigma(Y) \Delta X.$$

As before  $X(t)$  is normalized Brownian motion.  $Y(t)$  should turn out to be Markov with continuous paths and stationary transition probabilities. Argue this way:  $\mu(Y) \Delta t$  is a term approximately linear in  $t$ , but except for this term  $\Delta Y$  is of the order of  $\Delta X$ , hence  $Y(t)$  should be continuous. Further, assume  $Y(t)$  is measurable  $\mathcal{F}(X(\tau), \tau \leq t)$ . Then the distribution of  $\Delta Y$  depends only on  $Y(t)$ , (through  $\mu(Y)$  and  $\sigma(Y)$ ), and on  $\Delta X$ , which is independent of  $\mathcal{F}(Y(\tau), \tau \leq t)$  with distribution depending only on  $\Delta t$ .

Roughly, a process satisfying (16.9) is *locally Brownian*. Given that  $Y(t) = y$ , it behaves for the next short time interval as though it were a Brownian motion with drift  $\mu(y)$ , variance  $\sigma(y)$ . Therefore, we can think of constructing this kind of process by patching together various Brownian motions. Note, assuming  $\Delta X$  is independent of  $Y(t)$ ,

$$E(\Delta Y | Y(t) = y) = \mu(y) \Delta t,$$

$$E((\Delta Y)^2 | Y(t) = y) = \sigma^2(y)E(\Delta X)^2 + o(\Delta t) = \sigma^2(y) \Delta t + o(\Delta t).$$

Of course, the continuity condition is also satisfied,

$$P(|\Delta Y| > \epsilon | Y(t) = y) = o(\Delta t).$$

Define the truncated change in  $Y$  by

$$\Delta_\epsilon Y = \begin{cases} \Delta Y & \text{if } |\Delta Y| \leq \epsilon, \\ 0, & \text{otherwise.} \end{cases}$$

As a first approximation to the subject matter of this chapter, I will say that we are going to look at Markov processes  $Y(t)$  taking values in some interval

$F$ , with stationary transition probabilities  $p_t$  satisfying for every  $\epsilon > 0$ , and  $y$  in the interior of  $F$ ,

$$\begin{aligned} P(|\Delta Y| > \epsilon \mid Y(t) = y) &= o(\Delta t), \\ (16.10) \quad E(\Delta_\epsilon Y \mid Y(t) = y) &= \mu(y) \Delta t + o(\Delta t), \\ E((\Delta_\epsilon Y)^2 \mid Y(t) = y) &= \sigma^2(y) \Delta t + o(\Delta t), \end{aligned}$$

and having continuous sample paths. Conditions (16.10) are the infinitesimal conditions for the process. A Taylor expansion gives

**Proposition 16.11.** *Let  $f(x)$  be bounded with a continuous second derivative. If  $p_t(dy \mid x)$  satisfies (16.10), then  $(Sf)(x)$  exists for every point  $x$  in the interior of  $F$  and equals*

$$\frac{1}{2}\sigma^2(x) \frac{d^2 f(x)}{dx^2} + \mu(x) \frac{df(x)}{dx}.$$

Thus,

**Proposition 16.12.** *If the transition probabilities satisfy (16.10), and have densities  $p_t(y \mid x)$  with a continuous second derivative for  $x \in \text{int}(F)$ , then*

$$(16.13) \quad \frac{\partial}{\partial t} p_t(y \mid x) = \frac{1}{2}\sigma^2(x) \frac{\partial^2}{\partial x^2} p_t(y \mid x) + \mu(x) \frac{\partial}{\partial x} p_t(y \mid x), \quad x \in \text{int}(F).$$

*Proof.* This is the backwards equation.

**Problem 3.** Show by direct computation that the transition probabilities for the Ornstein-Uhlenbeck process satisfy

$$\frac{\partial}{\partial t} p_t(y \mid x) = \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2} p_t(y \mid x) - \alpha x \frac{\partial}{\partial x} p_t(y \mid x).$$

### 3. BROWNIAN MOTION WITH BOUNDARIES

For  $X(t)$  a locally Brownian process as in the last section, the infinitesimal operator  $S$  is defined for all interior points of  $F$  by 16.11. Of course, this completely defines  $S$  if  $F$  has only interior points. But if  $F$  has a closed boundary point, the definition of  $S$  at this point is not clear. This problem is connected with the question of what boundary conditions are needed to uniquely solve the backwards equation (16.13). To illuminate this problem a bit, we consider two examples of processes where  $F$  has a finite closed boundary point.

**Definition 16.14.** Use  $X_0(t)$  to denote normalized Brownian motion on  $R^{(1)}$ ,  $p_t^{(0)}(dy \mid x)$  are its transition probabilities.

The examples will be concerned with Brownian motion restricted to the interval  $F = [0, \infty)$ .

**Example 1. Brownian motion with an absorbing boundary.** Take  $F = [0, \infty)$ . The Brownian motion  $X(t)$  starting from  $x > 0$  with absorption at  $\{0\}$  is defined by

$$X(t) = X_0(\min(t, \tau_0^*)),$$

where  $X_0(t)$  is started from  $x$ .

It is not difficult to check that  $X(t)$  is Markov with stationary transition probabilities. To compute these rigorously is tricky. Let  $A \subset (0, \infty)$ ,  $A \in \mathcal{B}_1$ , and consider  $P_x(X_0(t) \in A, \tau_0^* < t)$ . The set  $\{X_0(t) \in A, \tau_0^* < t\}$  consists of all sample paths that pass out of  $(0, \infty)$  at least once and then come back in and get to  $A$  by time  $t$ . Let  $A^\perp$  be the reflection of  $A$  around the point  $x = 0$ . Argue that after hitting  $\{0\}$  at time  $\tau < t$  it is just as probable (by symmetry) to get to  $A^\perp$  by time  $t$  as it is to get to  $A$ , implying

$$(16.15) \quad P_x(X_0(t) \in A, \tau_0^* < t) = P_x(X_0(t) \in A^\perp, \tau_0^* < t).$$

This can be proven rigorously by approximating  $\tau_0^*$  by stopping times that take only a countable number of values. We assume its validity. Proceed by noting that  $\{X_0(t) \in A^\perp\} \subset \{\tau_0^* < t\}$  so that

$$\begin{aligned} P_x(X_0(t) \in A^\perp, \tau_0^* < t) &= P_x(X_0(t) \in A^\perp) \\ &= p_t^{(0)}(A^\perp | x). \end{aligned}$$

Now,

$$\begin{aligned} P_x(X(t) \in A) &= P_x(X_0(t) \in A, \tau_0^* \geq t) \\ &= P_x(X_0(t) \in A) - P_x(X_0(t) \in A, \tau_0^* < t) \\ &= p_t^{(0)}(A | x) - p_t^{(0)}(A^\perp | x). \end{aligned}$$

The density for  $p_t^{(0)}(A^\perp | x)$  is  $p_t^{(0)}(-y | x)$ . Thus

$$(16.16) \quad p_t(y | x) = p_t^{(0)}(y | x) - p_t^{(0)}(-y | x).$$

**Example 2. Brownian motion with a reflecting boundary.** Define the Brownian motion  $X(t)$  on  $F = [0, \infty)$  with a reflecting boundary at  $\{0\}$  to be

$$X(t) = |X_0(t)|,$$

where we start the motion from  $x > 0$ . What this definition does is to take all parts of the  $X_0(t)$  path below  $x = 0$  and reflect them in the  $x = 0$  axis getting the  $X(t)$  path.

**Proposition 16.17.**  $X(t)$  is Markov with stationary transition probability density

$$p_t(y | x) = p_t^{(0)}(y | x) + p_t^{(0)}(-y | x).$$

*Proof.* Take  $A \in \mathcal{B}_1([0, \infty))$ ,  $x \geq 0$ . Consider the probabilities

$$P(X(t + \tau) \in A \mid X_0(t) = x, X_0(s), s \leq t),$$

$$P(X(t + \tau) \in A \mid X_0(t) = -x, X_0(s), s \leq t).$$

Because  $X_0(t)$  is Markov, these reduce to

$$P(|X_0(t + \tau)| \in A \mid X_0(t) = x),$$

$$P(|X_0(t + \tau)| \in A \mid X_0(t) = -x).$$

These expressions are equal. Hence

$$\begin{aligned} P(X(t + \tau) \in A \mid X(t) = x, X(s), s \leq t) \\ &= E(P(X(t + \tau) \in A \mid |X_0(t)| = x, X_0(s), s \leq t) \mid X(s), s \leq t) \\ &= P(|X_0(t + \tau)| \in A \mid X_0(t) = x) \\ &= p_t^{(0)}(A \mid x) + p_t^{(0)}(A^\perp \mid x). \end{aligned}$$

In both examples,  $X(t)$  equals the Brownian motion  $X_0(t)$  until the particle reaches zero. Therefore, in both cases, for  $x > 0$  and  $f$  bounded and continuous on  $[0, \infty)$ ,

$$|E_x f(X(t)) - E_x f(X_0(t))| = o(t).$$

As expected, in the interior of  $F$ , then

$$(Sf)(x) = \frac{1}{2} \frac{d^2 f(x)}{dx^2}$$

for functions with continuous second derivatives. Assume that the limits  $f'(0+)$ ,  $f''(0+)$ , as  $x \downarrow 0$  of the first and second derivatives, exist. In the case of a reflecting boundary at zero, direct computation gives

$$E_0 f(X(t)) - f(0) = \sqrt{\frac{t}{2\pi}} f'(0+) + \frac{t}{2} f''(0+) + o(t).$$

Thus,  $(Sf)(0)$  does not exist unless  $f'(0+) = 0$ . If  $f'(0+) = 0$ , then  $(Sf)(0) = \frac{1}{2} f''(0+)$ , so not only is  $(Sf)(x)$  defined at  $x = 0$ , but it is also continuous there.

If  $\{0\}$  is absorbing, then for any  $f$ ,  $(Sf)(0) = 0$ . If we want  $(Sf)(x)$  to be continuous at zero, we must add the restriction  $f''(0+) = 0$ .

Does the backwards equation (16.13) have the transition probabilities of the process as its unique solution? Even if we add the restriction that we will consider only solutions which are densities of transition probabilities of Markov processes, the examples above show that the solution is not

unique. However, note that in the case of absorption

$$\lim_{x \downarrow 0} p_t(y | x) = 0$$

for all  $t, y > 0$ . Intuitively this makes sense, because the probability starting from  $x$  of being absorbed at zero before hitting the point  $y$  goes to one as  $x \rightarrow 0$ . For reflection, use the symmetry to verify that

$$\lim_{x \downarrow 0} \frac{\partial}{\partial x} p_t(y | x) = 0.$$

If either of the above boundary conditions are imposed on the backwards equation, it is possible to show that there is a unique solution which is a set of transition probability densities.

Reflection or absorption is not the only type of behavior possible at boundary points. Odd things can occur, and it was the occurrence of some of these eccentricities which first prompted Feller's investigation [56] and eventually led to a complete classification of boundary behavior.

### Problems

4. Show that the process  $X(t)$  defined on  $[0, 1]$  by folding over Brownian motion,

$$X(t) = |X_0(t) - 2n|$$

if  $|X_0(t) - 2n| \leq 1$ , is a Markov process with stationary transition probabilities such that

$$\left. \frac{\partial}{\partial x} p_t(y | x) \right|_{0+} = \left. \frac{\partial}{\partial x} p_t(y | x) \right|_{1-} = 0.$$

Evaluate  $(Sf)(x)$  for  $x \in (0, 1)$ . For what functions does  $(Sf)(x)$  exist at the endpoints? [This process is called Brownian motion with two reflecting boundaries.]

5. For Brownian motion on  $[0, \infty)$ , either absorbing or reflecting, evaluate the density  $r_\lambda(y | x)$  of the resolvent  $R_\lambda(dy | x)$ . For  $\mathcal{C}$  the class of all bounded continuous functions on  $[0, \infty)$ , show that

a) For absorbing Brownian motion,

$$\mathcal{D}(S, \mathcal{C}) = \{f \in \mathcal{C}; f'' \in \mathcal{C}, f''(0+) = 0\}.$$

b) For reflecting Brownian motion,

$$\mathcal{D}(S, \mathcal{C}) = \{f \in \mathcal{C}; f'' \in \mathcal{C}, f'(0+) = 0\}.$$

[See Problems 10 and 11, Chapter 15. Note that in (a),  $R_\lambda(dy | 0)$  assigns all of its mass to  $\{0\}$ .]



## 4. FELLER PROCESSES

The previous definitions and examples raise a host of interesting and difficult questions. For example:

1) Given the form of  $S$ , that is, given  $\sigma^2(x)$  and  $\mu(x)$  defined on  $\text{int}(F)$ , and certain boundary conditions. Does there exist a unique Markov process with continuous paths having  $S$  as its infinitesimal operator and exhibiting the desired boundary behavior?

2) If the answer to (1) is Yes, do the transition probabilities have a density  $p_t(y|x)$ ? Are these densities smooth enough in  $x$  so that they are a solution of the backwards equations? Do the backwards equations have a unique solution?

The approach that will be followed is similar to that of the previous chapter. Investigate first the properties of the class of processes we want to construct—try to simplify their structure as far as possible. Then work backwards from the given infinitesimal conditions to a process of the desired type.

To begin, assume the following.

**Assumption 16.18(a).**  $F$  is an interval closed, open, or half of each, finite or infinite.  $X(t)$  is a Markov process with stationary transition probabilities such that starting from any point  $x \in F$ , all sample paths are continuous.

The next step would be to prove that 16.18(a) implies that the strong Markov property holds. *This is not true.* Consider the following counter example: Let  $F = \mathbb{R}^{(1)}$ . Starting from any  $x \neq 0$ ,  $X(t)$  is Brownian motion starting from that point. Starting from  $x = 0$ ,  $X(t) \equiv 0$ . Then, for  $x \neq 0$ ,

$$P_x(X(t + \tau_0^*) \in B) = P_0(X_0(t) \in B).$$

But if  $X(t)$  were strong Markov, since  $\tau_0^*$  is a Markov time,

$$P_x(X(t + \tau_0^*) \in B) = P_{X(\tau_0^*)}(X(t) \in B) = \chi_B(0).$$

The pathology in this example is that starting from the point  $x = 0$  gives distributions drastically different from those obtained by starting from any point  $x \neq 0$ . When you start going through the proof of the strong Markov property, you find that it is exactly this large change in the distribution of the process when the initial conditions are changed only slightly that needs to be avoided. This recalls the concept of stability introduced in Chapter 8.

**Definition 16.19.** Call the transition probabilities stable if for any sequence of initial distributions  $\pi_n(\cdot) \xrightarrow{\mathcal{D}} \pi(\cdot)$ , the corresponding probabilities satisfy

$$P_{\pi_n}(X(t) \in \cdot) \xrightarrow{\mathcal{D}} P_{\pi}(X(t) \in \cdot)$$

for any  $t > 0$ . Equivalently, for all  $\varphi(x)$  continuous and bounded on  $F$ ,  $E_x\varphi(X(t))$  is continuous on  $F$ .

Assume, in addition to 16.18(a),

**Assumption 16.18(b).** *The transition probabilities of  $X(t)$  are stable.*

**Definition 16.20.** *A process satisfying 16.18(a),(b) will be called a Feller process.*

Now we can carry on.

**Theorem 16.21.** *A Feller process has the strong Markov property.*

*Proof.* Let  $\varphi(x)$  be bounded and continuous on  $F$ , then  $\tilde{\varphi}(x) = E_x \varphi(X(t))$  is likewise. Let  $\tau^*$  be a Markov time,  $\tau_n^*$  a sequence of Markov times such that  $\tau_n^* \downarrow \tau^*$  and  $\tau_n^*$  takes on only a countable number of values. From (15.20), for  $C \in \mathcal{F}(X(t), t \leq \tau^*)$ ,

$$\int_C \varphi(X(t + \tau_n^*)) dP_x = \int_C \tilde{\varphi}(X(\tau_n^*)) dP_x.$$

The path continuity and the continuity of  $\varphi, \tilde{\varphi}$  give

$$\begin{aligned} \varphi(X(t + \tau_n^*)) &\xrightarrow{\text{a.s.}} \varphi(X(t + \tau^*)), \\ \tilde{\varphi}(X(\tau_n^*)) &\xrightarrow{\text{a.s.}} \tilde{\varphi}(X(\tau^*)). \end{aligned}$$

Thus, for  $\varphi$  continuous,

$$E_x(\varphi(X(t + \tau^*)) | X(\tau), \tau \leq \tau^*) = E_{X(\tau^*)}(\varphi(X(t))).$$

The continuous functions separate, thus, for any  $B \in \mathcal{B}_1(F)$ ,

$$P_x(X(t + \tau^*) \in B | X(\tau), \tau \leq \tau^*) = P_{X(\tau^*)}(X(t) \in B).$$

To extend this, let  $\varphi(x_1, \dots, x_k)$  on  $F^{(k)}$  equal a product  $\varphi_1(x_1) \cdots \varphi_k(x_k)$ , where  $\varphi_1, \dots, \varphi_k$  are bounded and continuous on  $F$ . It is easy to check that

$$E_x \varphi(X(t_k), \dots, X(t_1)) = \tilde{\varphi}(x)$$

is continuous in  $x$ . By the same methods as in (15.20), get

$$\int_C \varphi(X(t_k + \tau_n^*), \dots, X(t_1 + \tau_n^*)) dP_x = \int_C \tilde{\varphi}(X(\tau_n^*)) dP_x,$$

conclude that

$$\int_C \varphi(X(t_k + \tau^*), \dots, X(t_1 + \tau^*)) dP_x = \int_C \tilde{\varphi}(X(\tau^*)) dP_x,$$

and now use the fact that products of bounded continuous functions separate probabilities on  $\mathcal{B}_k(F)$ .

Of course, this raises the question of how much stronger a restriction stability is than the strong Markov property. The answer is—Not much!

To go the other way, it is necessary that the state space have something like an indecomposability property—that every point of  $F$  can be reached from every interior point.

**Definition 16.22.** The process is called *regular* if for every  $x \in \text{int}(F)$  and  $y \in F$ ,

$$P_x(\tau_y^* < \infty) > 0.$$

**Theorem 16.23.** If  $X(t)$  is regular and strong Markov, then its transition probabilities are stable.

*Proof.* Let  $x < y < z$ ,  $x \in \text{int}(F)$ . Define  $\bar{t}_y = \min(\tau_y^*, s)$ ,  $\bar{t}_z = \min(\tau_z^*, s)$ , for  $s > 0$  such that  $P_x(\tau_y^* \leq s) > 0$ . These are Markov times. Take  $\varphi(x)$  bounded and continuous,  $\tilde{\varphi}(x) = E_x\varphi(X(t))$ . By the strong Markov property,

$$\begin{aligned} E_x\varphi(X(t + \bar{t}_y)) &= E_x[E_x(\varphi(X(t + \bar{t}_y)) \mid X(\tau), \tau \leq \bar{t}_y)] \\ &= E_x\tilde{\varphi}(X(\bar{t}_y)), \end{aligned}$$

$$E_x\varphi(X(t + \bar{t}_z)) = E_x\tilde{\varphi}(X(\bar{t}_z)).$$

Suppose that on the set  $\{\tau_y^* < \infty\}$ ,  $\tau_y^* \uparrow \tau_z^*$  a.s.  $P_x$  as  $y \uparrow z$ , implying  $\bar{t}_y \uparrow \bar{t}_z$  a.s. So  $E_x\varphi(X(t + \bar{t}_y)) \rightarrow E_x\varphi(X(t + \bar{t}_z))$ . The right-hand sides of the above equations are

$$\tilde{\varphi}(y)P_x(\tau_y^* \leq s) + \int_{\{\tau_y^* > s\}} \tilde{\varphi}(X(s)) dP_x$$

and

$$\tilde{\varphi}(z)P_x(\tau_z^* \leq s) + \int_{\{\tau_z^* > s\}} \tilde{\varphi}(X(s)) dP_x.$$

For  $s$  a continuity point of  $P_x(\tau_z^* \leq s)$ ,  $P_x(\tau_y^* \leq s) \rightarrow P_x(\tau_z^* \leq s)$ , and the sets  $\{\tau_y^* > s\} \uparrow \{\tau_z^* > s\}$ . The conclusion is that  $\tilde{\varphi}(y) \rightarrow \tilde{\varphi}(z)$ .

The final part is to get  $\tau_y^* \uparrow \tau_z^*$  as  $y \uparrow z$ . Let  $\tau_z^*(\omega) < \infty$  and denote  $\tau = \lim \tau_y^*$ , as  $y \uparrow z$ . On the set  $\{\tau_y^* < \infty\}$ ,  $X(\tau_y^*) = y$ . By the path continuity,  $X(\tau) = z$ , so  $\tau = \tau_z^*$ . By varying  $x$  to the left and right of points, 16.23 results.

From here on, we work only with regular processes.

**Problem 6.** Let  $\mathcal{C}$  be the class of all bounded continuous functions on  $F$ . For a Feller process show that if  $f \in \mathcal{C}$ , then  $R_\lambda f$  is in  $\mathcal{C}$ . Deduce that  $\mathcal{D}(S, \mathcal{C}) \subset \mathcal{D}(S, \mathcal{G})$ . Prove that there is at most one Feller process corresponding to a given  $S$ ,  $\mathcal{D}(S, \mathcal{C})$ . (See Theorem 15.53.)

## 5. THE NATURAL SCALE

For pure jump processes, the structure was decomposable into a space structure, governed by a discrete time Markov chain and a time rate which determined how fast the particle moved through the paths of the space

structure. Regular Feller processes can be decomposed in a very similar way. The idea is clearer when it is stated a bit more generally. Look at a path-continuous Markov process with stable stationary transition probabilities taking values in  $n$ -dimensional space,  $\mathbf{X}(t) \in R^{(n)}$ . Consider a set  $B \in \mathcal{B}_n$ ; let  $\mathbf{t}^*(B)$  be the first exit time from  $B$ , then using the nontrivial fact that  $\mathbf{t}^*(B)$  is measurable, define probabilities on the Borel subsets of the boundary of  $B$  by

$$Q_x(C) = P_x(\mathbf{X}(\mathbf{t}^*(B)) \in C), \quad C \subset bd(B).$$

The  $Q_x(C)$  are called exit distributions and specify the location of  $\mathbf{X}(t)$  upon first leaving  $B$ . Suppose two such processes have the same exit distributions for all open sets. Then we can prove that under very general conditions they differ from each other only by a random change of time scale [10]. Thus the exit distributions characterize the space structure of the process. To have the exit distributions make sense, it is convenient to know that the particle does exit a.s. from the set in question. Actually, we want and get much more than this.

**Theorem 16.24.** *Let  $J$  be a finite open interval such that  $\bar{J} \subset F$ . Then*

$$\sup_{x \in J} E_x \mathbf{t}^*(J) < \infty.$$

*Proof.* First we need

**Lemma 16.25.** *Let  $\sup_{x \in J} P_x(\mathbf{t}^*(J) > t) = \alpha < 1$  for some  $t > 0$ . Then*

$$E_x \mathbf{t}^*(J) \leq \frac{t}{1 - \alpha}.$$

*Proof.* Let  $\alpha_n = \sup_{x \in J} P_x(\mathbf{t}^*(J) > nt)$ . Write

$$P_x(\mathbf{t}^*(J) > (n+1)t) \leq \alpha_n P_x(\mathbf{t}^*(J) > (n+1)t \mid \mathbf{t}^*(J) > nt).$$

Let  $\mathbf{t}_n^*(J)$  be the first exit time from  $J$ , starting from time  $nt$ . Then

$$P_x(\mathbf{t}^*(J) > (n+1)t) \leq \alpha_n P_x(\mathbf{t}_n^*(J) > t \mid \mathbf{t}^*(J) > nt).$$

Since  $\{\mathbf{t}^*(J) > nt\} \in \mathcal{F}(\mathbf{X}(\tau), \tau \leq nt)$ , then

$$P_x(\mathbf{t}^*(J) > (n+1)t) \leq \alpha \alpha_n.$$

Hence  $\alpha_n \leq \alpha^n$ , and

$$\begin{aligned} E_x \mathbf{t}^*(J) &= \int_0^\infty P_x(\mathbf{t}^*(J) > \tau) d\tau \\ &= \sum_{n=0}^\infty \int_{nt}^{(n+1)t} P_x(\mathbf{t}^*(J) > \tau) d\tau \\ &\leq t \sum_{n=0}^\infty \alpha^n = \frac{t}{1 - \alpha}. \end{aligned}$$

To finish the proof of 16.24, let  $J = [a, b]$ , pick  $y \in J$ . By regularity, there exists a  $t$  and  $\alpha$  such that  $P_y(t_a^* > t) \leq \alpha < 1$ ,  $P_y(t_b^* > t) \leq \alpha < 1$ . For  $y \leq x < b$ ,

$$P_x(t^*(J) > t) \leq P_x(t_b^* > t) \leq P_y(t_b^* > t),$$

and for  $a < x \leq y$ ,

$$P_x(t^*(J) > t) \leq P_x(t_a^* > t) \leq P_y(t_a^* > t).$$

Apply the lemma now to deduce the theorem.

**Remark.** Note that the lemma has wider applicability than the theorem. Actually, it holds for all intervals  $J$ , finite or infinite.

In one dimension the relevant exit probabilities are:

**Definition 16.26.** For any open interval  $J = (a, b)$  such that  $P_x(t^*(J) < \infty) = 1$ ,  $x \in J$ , define

$$p^+(x, J) = P_x(X(t^*(J)) = b),$$

$$p^-(x, J) = 1 - p^+(x, J).$$

**Theorem 16.27.** There exists a continuous, strictly increasing function  $u(x)$  on  $F$ , unique up to a linear transformation, such that for  $J \subset F$ ,  $J = (a, b)$ ,

$$(16.28) \quad p^+(x, J) = \frac{u(x) - u(a)}{u(b) - u(a)}.$$

*Proof.* For  $J \subset I$ , note that exiting right from  $I$  starting from  $x \in J$  can occur in two ways:

- i) Exit right from  $J$  starting from  $x$ , then exit right from  $I$  starting from  $b$ .
- ii) Exit left from  $J$  starting from  $x$ , then exit right from  $I$  starting from  $a$ .

Use the strong Markov property to conclude that for  $x \in J$ ,

$$(16.29) \quad p^+(x, I) = p^+(x, J)p^+(b, I) + p^-(x, J)p^+(a, I)$$

or

$$p^+(x, J) = \frac{p^+(x, I) - p^+(a, I)}{p^+(b, I) - p^+(a, I)}.$$

If  $F$  were bounded and closed, then we could take  $u(x) = p^+(x, \text{int}(F))$  and satisfy the theorem. In general, we have to use extension. Take  $I_0$  to be a bounded open interval, such that if  $F$  includes any one of its endpoints, then  $I_0$  includes that endpoint. Otherwise  $I_0$  is arbitrary. Define  $u(x)$  on  $I_0 = (x_1, x_2)$  as  $p^+(x, I_0)$ . By the equation above, for  $I_0 \subset I_1$ ,  $x \in I_0$ ,

$$(16.30) \quad u(x) = \frac{p^+(x, I_1) - p^+(x_1, I_1)}{p^+(x_2, I_1) - p^+(x_1, I_1)}.$$

Define an extension  $u^{(1)}(x)$  on  $I_1$  by the right-hand side of (16.30). Suppose another interval  $I_2$  is used to get an extension, say  $I_1 \subset I_2$ . Then for  $x \in I_2$ , we would have

$$u^{(2)}(x) = \frac{p^+(x, I_2) - p^+(x_1, I_2)}{p^+(x_2, I_2) - p^+(x_1, I_2)}.$$

For  $x \in I_1$ , (16.29) gives

$$p^+(x, I_2) = c_1 p^+(x, I_1) + c_2.$$

Substitute this into (16.30) to conclude that  $u^{(1)}(x) = u^{(2)}(x)$  on  $I_1$ . Thus the extensions are unique. Continuing this way, we can define  $u(x)$  on  $\text{int}(F)$  so that (16.28) is satisfied.

It is increasing; otherwise there exists a finite open  $J$ ,  $J \subset F$ , and  $x \in J$ , such that  $p^+(x, J) = 0$ . This contradicts regularity. Extend  $u$ , by taking limits, to endpoints of  $F$  included in  $F$ . Now let  $J_n$  be open,  $J_n \uparrow J = (a, b)$ . I assert that  $t^*(J_n) \uparrow t^*(J)$ . Because  $t^*(J_n) \leq t^*(J)$ , by monotonicity  $t^* = \lim_n t^*(J_n)$  exists, and by continuity  $X(t^*) = a$  or  $b$ . For any  $\epsilon > 0$  and  $n$  sufficiently large,

$$p^+(x, J_n) = P_x(X(t^*(J_n)) \geq b - \epsilon).$$

Since  $X(t^*(J_n)) \rightarrow X(t^*(J))$  a.s., taking limits of the above equation gives

$$p^+(x, J_n) \rightarrow p^+(x, J).$$

By taking either  $a_n \downarrow a$  or  $b_n \uparrow b$  we can establish the continuity. The fact that  $u(x)$  is unique up to a linear transformation follows from (16.28).

We will say that a process is on its natural scale if it has the same exit distributions as Brownian motion. From (13.3),

**Definition 16.31.** A process is said to be on its natural scale if for every  $J = (a, b)$ ,  $\bar{J} \subset F$ ,

$$p^+(x, J) = \frac{x - a}{b - a},$$

that is, if  $u(x) = x$  satisfies (16.28).

The distinguishing feature of the space structure of normalized Brownian motion is that it is *driftless*. There is as much tendency to move to the right as to the left. More formally, if  $J$  is any finite interval and  $x_0$  its midpoint, then for normalized motion, by symmetry,  $p^+(x_0, J) = \frac{1}{2}$ . We generalize this to

**Proposition 16.32.** A process is on its natural scale if and only if for every finite open  $J$ ,  $\bar{J} \subset F$ ,  $x_0$  the midpoint of  $J$ ,

$$p^+(x_0, J) = \frac{1}{2}.$$

*Proof.* Consider  $n$  points equally spaced in  $J$ ,

$$a = x_1 < \cdots < x_n = b.$$

Starting from  $x_k$ , the particle next hits  $x_{k-1}$  or  $x_{k+1}$  with equal probability, so  $p^+(x_k, (x_{k-1}, x_{k+1})) = \frac{1}{2}$ . Therefore, the particle behaves like a symmetric random walk on the points of the partition. From Chapter 7, Section 10,

$$p^+(x_k, J) = \frac{x_k - a}{b - a}.$$

The continuity established in Theorem 16.27 completes the proof of 16.32.

Let the state space undergo the transformation  $\tilde{x} = u(x)$ . Equivalently, consider the process

$$(16.33) \quad \tilde{X}(t) = u(X(t)).$$

If  $X(t)$  is a regular Feller process, then so is  $\tilde{X}(t)$ . The importance of this transformation is:

**Proposition 16.34.**  $\tilde{X}(t)$  is on its natural scale.

*Proof.* Let  $\tilde{J} = (\tilde{a}, \tilde{b})$ ,  $\tilde{a} = u(a)$ ,  $\tilde{b} = u(b)$ ,  $J = (a, b)$ . For the  $\tilde{X}(t)$  process, with  $\tilde{x} = u(x)$ ,

$$\begin{aligned} p^+(\tilde{x}, \tilde{J}) &= P_{\tilde{x}}(\tilde{X}(\tau^*(\tilde{J})) = \tilde{b}) \\ &= P_x(X(\tau^*(J)) = b) \\ &= \frac{u(x) - u(a)}{u(b) - u(a)} \\ &= \frac{\tilde{x} - \tilde{a}}{\tilde{b} - \tilde{a}}. \end{aligned}$$

For any regular Feller process then, a simple space transformation gives another regular Feller process having the same space structure as normalized Brownian motion. *Therefore, we restrict attention henceforth to this type and examine the time flow.*

**Remark.** The reduction to natural scale derived here by using the transformation  $\tilde{x} = u(x)$  does not generalize to Feller processes in two or more dimensions. Unfortunately, then, a good deal of the theory that follows just does not generalize to higher dimensions.

## 6. SPEED MEASURE

The functions  $m(x, J) = E_x \tau^*(J)$  for open intervals  $J$ , determine how fast the process moves through its paths. There is a measure  $m(dx)$  closely

associated with these functions. Define, for  $J = (a, b)$  finite,

$$(16.35) \quad G_J(x, y) = \begin{cases} \frac{2(x-a)(b-y)}{b-a}, & x, y \in J, \quad x \leq y, \\ \frac{2(y-a)(b-x)}{b-a}, & x, y \in J, \quad x \geq y, \\ 0, & \text{otherwise.} \end{cases}$$

Then,

**Theorem 16.36.** *Let  $X(t)$  be on its natural scale. Then there is a unique measure  $m(dx)$  defined on  $\mathcal{B}_1(\text{int } F)$ ,  $m(B) < \infty$  for  $B$  bounded,  $\bar{B} \subset \text{int}(F)$ , such that for finite open  $J, \bar{J} \subset F$ ,*

$$m(x, J) = \int G_J(x, y) m(dy).$$

*Proof.* The proof will provide some justification for the following terminology.

**Definition 16.37.**  $m(dx)$  is called the speed measure for the process.

Consider  $(a, b)$  partitioned by points  $a = x_0 < x_1 < \dots < x_n = b$ , where  $x_k = a + k\delta$ . Define  $J_k = (x_{k-1}, x_{k+1})$ . Note that  $m(x_k, J_k)$  gives some quantitative indication of how fast the particle is traveling in the vicinity of  $x_k$ . Consider the process only at the exit times from one of the  $J_k$  intervals. This is a symmetric random walk  $Z_0, Z_1, \dots$  moving on the points  $x_0, \dots, x_n$ . Let  $n(x_j; x_k)$  be the expected number of times the random walk visits the point  $x_k$ , starting from  $x_j$  before hitting  $x_0$  or  $x_n$ . Then

$$(16.38) \quad m(x_j, J) = \sum_{k=1}^{n-1} n(x_j; x_k) m(x_k, J_k).$$

This formula is not immediately obvious, because  $t^*(I)$  and  $X(t^*(I))$  are not, in general, independent. Use this argument: Let  $t_N^*$  be the time taken for the transition  $Z_N \rightarrow Z_{N+1}$ . For  $x \in \{x_0, \dots, x_n\}$ ,

$$E_x t_N^* = E_x(E_x(t_N^* | Z_N)) = \sum_{k=1}^{n-1} P_x(Z_N = x_k) m(x_k, J_k).$$

Sum over  $N$ , noting that

$$n(x_j; x_k) = E_{x_j} \left( \sum_{N=0}^{\infty} \chi_{\{x_k\}}(Z_N) \right) = \sum_{N=0}^{\infty} P_{x_j}(Z_N = x_k).$$

This function was evaluated in Problem 14, Chapter 7, with the result,

$$n(x_j; x_k) = \frac{1}{\delta} G_J(x_j, x_k).$$



Defining  $\hat{m}(dx)$  as a measure that gives mass  $m(x_k, J_k)/\delta$  to the point  $x_k$ , we get

$$m(x_j, J) = \int G_J(x_j, y) \hat{m}(dy).$$

Now to get  $m(dx)$  defined on all  $\text{int}(F)$ . Partition  $F$  by successive refinements  $\mathcal{F}^{(n)}$  having points a distance  $\delta_n$  apart, with  $\delta_n \rightarrow 0$ . Define the measure  $m_n$  as assigning mass

$$m(x_k, (x_{k-1}, x_{k+1}))/\delta_n$$

to all points  $x_k \in \mathcal{F}^{(n)}$  which are not endpoints of the partition. For  $a, b, x_j \in \mathcal{F}^{(n)}, J = (a, b), J \subset F$ ,

$$(16.39) \quad m(x_j, J) = \int G_J(x_j, y) m_n(dy).$$

For any finite interval  $I$  such that  $I \subset \text{int}(F)$ , (16.39) implies that  $\overline{\lim}_n m_n(I) < \infty$ . Use this fact to conclude that there exists a subsequence  $m_{n'} \xrightarrow{w} m$ , where  $m$  is a measure on  $\mathcal{B}_1(\text{int } F)$  (see 10.5). Furthermore, for any  $J = (a, b)$ , and  $x \in J$ , where  $a, b, x$  are in  $\bigcup_n \mathcal{F}^{(n)}$ ,

$$(16.40) \quad m(x, J) = \int G_J(x, y) m(dy).$$

For any arbitrary finite open interval  $J, J \subset F$ , take  $J_1 \subset J$  where  $J_1$  has endpoints in  $\bigcup_n \mathcal{F}^{(n)}$  and pass to the limit as  $J_1 \uparrow J$ , to get (16.40) holding for  $J$  and any  $x \in \bigcup_n \mathcal{F}^{(n)}$ . To extend this to arbitrary  $x$ , we introduce an identity: If  $I \subset J, I = (a, b), x \in I$ , then the strong Markov property gives

$$(16.41) \quad m(x, J) = m(x, I) + p^+(x, I)m(b, J) + p^-(x, I)m(a, J).$$

Take  $J$  finite and open,  $J \subset F, x \in J, y < z < x$  and  $y, z \in \bigcup \mathcal{F}^{(n)} \cap J$ . Use (16.41) to write

$$(16.42) \quad m(z, J) = m(z, (y, x)) + p^+(z, (y, x))m(x, J) + p^-(z, (y, x))m(y, J).$$

Take  $z \uparrow x$ . By (16.40),  $m(z, (y, x)) \rightarrow 0$ , so

$$\lim_{z \uparrow x} m(z, J) = m(x, J).$$

Since the integral  $\int G_J(x, y)m(dy)$  is continuous in  $x$ , then (16.40) holds for all  $x \in J$ . But now the validity of (16.40) and the fact that the set of functions  $\{G_J(x, y)\}, a, b, x \in \text{int}(F)$ , are separating on  $\text{int}(F)$  imply that  $m(dy)$  is unique, and the theorem follows.

One question left open is the assignment of mass to closed endpoints of  $F$ . This we defer to Section 7.

**Problems**

7. If  $X(t)$  is not on its natural scale, show that  $m(dx)$  can still be defined by

$$m(x, J) = \int G_J(x, y) m(dy)$$

by using the definition

$$G_J(x, y) = \begin{cases} \frac{2(u(x) - u(a))(u(b) - u(y))}{u(b) - u(a)}, & x, y \in J, \quad x \leq y, \\ \frac{2(u(y) - u(a))(u(b) - u(x))}{u(b) - u(a)}, & x, y \in J, \quad x \geq y, \\ 0, & \text{otherwise.} \end{cases}$$

8. For  $X(t)$  Brownian motion with zero drift, show that for  $J = (a, b)$ ,

$$E_x \tau^*(J) = \frac{(x - a)(b - x)}{\sigma^2}.$$

Use this to prove

$$m(dy) = \frac{1}{\sigma^2} dy.$$

[Another way to see the form of  $m(dy)$  for Brownian motion is this:  $m_n(dy)$  assigns equal masses to points equally spaced. The only possible limit measure of measures of this type is easily seen to be a constant multiple  $c dy$  of Lebesgue measure. If  $c_0$  is the constant for normalized Brownian motion, a scale transformation of  $X(t)$  shows that  $c = c_0/\sigma^2$ .]

9. For  $x \in \text{int}(F)$  and  $J_n$  open neighborhoods of  $x$  such that  $J_n \downarrow \{x\}$ , show that  $\tau^*(J_n) \xrightarrow{\text{a.s.}} 0$ . Deduce that  $\tau^*(\{x\}) = 0$  a.s.  $P_x$ .

10. For  $f(x)$  a bounded continuous function on  $F$ ,  $J$  any finite open interval such that  $\bar{J} \subset F$ , prove that

$$E_x \left[ \int_0^{\tau^*(J)} f(X(t)) dt \right] = \int_J G_J(x, y) f(y) m(dy).$$

[Use the approximation by random walks on partitions.]

**7. BOUNDARIES**

This section sets up classifications that summarize the behavior at the boundary points of  $F$ . If  $F$  is infinite in any direction, say to the right, call  $+\infty$  the boundary point on the right; similarly for  $-\infty$ .

For a process on its natural scale the speed measure  $m(dx)$  defined on  $\text{int}(F)$  will to a large extent determine the behavior of the process at the boundaries of  $F$ . For example, we would expect that knowing the speed

measure in the vicinity of a boundary point  $b$  of  $F$  would tell us whether the process would ever hit  $b$ , hence whether  $b$  was an open or closed endpoint of  $F$ . Here, by open endpoint, we mean  $b \notin F$ , by closed,  $b \in F$ . In fact, we have

**Proposition 16.43.** *If  $b$  is a finite endpoint of  $F$ , then  $b \notin F$  if and only if for  $J \subset F$  any nonempty open neighborhood with  $b$  as endpoint,*

$$\int_J |b - y| m(dy) = \infty.$$

**Remark.** A closed endpoint of  $F$  is sometimes called *accessible*, for obvious reasons.

*Proof.* Assume  $b$  is a right endpoint. If  $b \in F$ , then there exists a  $t$  such that for  $c \in \text{int}(F)$ ,  $P_c(\tau_b^* > t) = \alpha < 1$ . Let  $J = (c, b)$ . Then  $P_x(\tau^*(J) > t) \leq \alpha$ , all  $x \in J$ . Use Lemma 16.25 to get  $m(x, J) < \infty$ , all  $x \in J$ . For  $b \in F$ ,  $J \subset F$ ; hence 16.36 holds:

$$\begin{aligned} m(x, J) &= \int_J G_J(x, y) m(dy) \\ &\geq \frac{2(x - c)}{b - c} \int_{(x, b)} |b - y| m(dy). \end{aligned}$$

Therefore the integral over  $(x, b)$  is finite. Conversely, if the integral is finite for one  $c \in \text{int}(F)$ , then as  $z \uparrow b$ , if  $b \notin F$ , and  $x > c$ ,

$$E_x \tau_c^* = \lim m(x, (c, z)) = \int_J G_J(x, y) m(dy), \quad J = (c, b).$$

The left-hand side of this is nondecreasing as  $x \uparrow b$ . But the integral equals

$$\frac{2(b - x)}{b - c} \int_{(c, x)} (y - c) m(dy) + \frac{2(x - c)}{b - c} \int_{[x, b)} (b - y) m(dy),$$

which goes to zero as  $x \uparrow b$ .

For open boundary points  $b$ , there is two-way classification.

**Definition 16.44.** Let  $x \in \text{int}(F)$ ,  $y \in \text{int}(F)$ ,  $y \rightarrow b$  monotonically. Call  $b$

natural if for all  $t > 0$ ,  $\lim_{y \rightarrow b} P_y(\tau_x^* < t) = 0$ ;

entrance if there is a  $t > 0$  such that  $\lim_{y \rightarrow b} P_y(\tau_x^* < t) > 0$ .

A natural boundary behaves like the points at  $\infty$  for Brownian motion. It takes the particle a long time to get close to such a boundary point and then a long time to get away from it. An entrance boundary has the odd property that it takes a long time for the particle to get out to it but not to get away from it.

**Proposition 16.45.** *Let  $b$  be an open boundary point.*

- a) *If  $b$  is finite, it must be natural.*  
 b) *If  $b$  is infinite, it is natural if and only if for  $J \subset F$  any open interval with  $b$  as an endpoint,*

$$\int_J |y| m(dy) = \infty.$$

*Proof.* Take  $b$  finite and right-hand, say. If  $b$  is an entrance point, then for  $J = (c, b)$ , use Lemma 16.25 to get  $m(x, J) \leq M < \infty$  for all  $x \in J$ . This implies, if we take  $J_1 \uparrow J$  and use the monotone convergence theorem, that

$$\int_J (b - y)m(dy) < \infty.$$

This is impossible, by 16.43, since  $b$  is an open endpoint. For  $b = \infty$ , check that

$$\lim_{\uparrow \infty} m(x, (a, c)) = 2 \int_a^x (y - a) dm + 2(x - a) \int_x^\infty dm.$$

If  $b = \infty$  is entrance, then there is an  $a$  such that for  $J = (a, \infty)$

$$P_x(\tau^*(J) > t) = P_x(\tau_a^* > t) \leq 1 - \alpha < 1$$

for all  $x \in J$ . Use Lemma 16.25 to get

$$\int_a^x (y - a) dm + (x - a) \int_x^\infty dm \leq M < \infty, \quad x \in J.$$

Taking  $x \rightarrow \infty$  proves that  $\int_J |y| dm < \infty$ . Conversely, if the integral is finite for  $J = (a, \infty)$ , then there exists an  $M < \infty$  such that for all  $a < x < c < \infty$ ,  $m(x, (a, c)) \leq M$ . Assume  $b = \infty$  is not an entrance boundary. For any  $0 < \epsilon < \frac{1}{4}$ , take  $c, x$  such that  $a < x < c$  and

$$P_x(\tau_a^* \leq 2M) \leq \epsilon, \quad P_x(\tau_c^* \leq 2M) \leq \epsilon.$$

Then

$$m(x, (a, c)) \geq 2MP_x(\tau^*((a, c)) \geq 2M) \geq 2M(1 - 2\epsilon) > M,$$

and this completes the proof.

There is also a two-way classification of closed boundary points  $b$ . This is connected with the assignment of speed measure  $m(\{b\})$  on these points. Say  $b$  is a closed left-hand endpoint. We define a function  $G_J(x, y)$  for all intervals of the form  $J = [b, c)$ , and  $x, y \in J$  as follows: Let  $\overset{\perp}{J}$  be the reflection of  $J$  around  $b$ , that is,  $\overset{\perp}{J} = (b - (c - b), b]$  and  $\overset{\perp}{y}$  the reflection of  $y$  around  $b$ ,  $\overset{\perp}{y} = b - (y - b)$ . Then define

$$(16.46) \quad G_J(x, y) = G_{J \cup \overset{\perp}{J}}(x, y) + G_{J \cup \overset{\perp}{J}}(x, \overset{\perp}{y}),$$

where  $G_{J \cup J^\perp}$  is defined by (16.35). This definition leads to an extension of (16.40).

**Theorem 16.47.** *It is possible to define  $m(\{b\})$  so that for all finite intervals  $J = [b, c)$ ,  $J \subset F$ ,*

$$m(x, J) = \int_J G_J(x, y) m(dy).$$

**Remark.** Note that if  $\int_{(b,c)} m(dy) = \infty$ , then no matter how  $m(\{b\})$  is assigned,  $m(x, J) = \infty$ . This leads to

**Definition 16.48.** *Let  $b$  be a closed boundary point,  $J \subset F$  any finite open interval with  $b$  as one endpoint such that  $J$  is smaller than  $F$ . Call  $b$*

$$\begin{array}{ll} \text{regular} & \text{if } m(J) < \infty, \\ \text{exit} & \text{if } m(J) = \infty. \end{array}$$

*Proof of 16.47.* The meaning of the definition of  $G_J$  will become more intuitive in the course of this proof. Partition any finite interval  $[b, c)$  by equally spaced points  $b = x_0 < x_1 < \cdots < x_n = c$ , a distance  $\delta$  apart. Define  $J_k = (x_{k-1}, x_{k+1})$ ,  $k = 1, \dots, n-1$ ,  $J_0 = [x_0, x_1)$ . By exactly the same reasoning as in deriving 16.38,

$$m(x_j, J) = \sum_{k=0}^{n-1} n^{(r)}(x_j; x_k) m(x_k, J_k),$$

where  $n^{(r)}(x_j; x_k)$  is the expected number of visits to  $x_k$  of a random walk  $Z_0, Z_1, \dots$  on  $x_0, \dots, x_n$  starting from  $x_j$ , with reflection at  $x_0$  and absorption at  $x_n$ . Construct a new state space  $x_{-n}, \dots, x_{-1}, x_0, \dots, x_n$ , where  $x_{-k} = \overset{\perp}{x}_k$ , and consider a symmetric random walk  $\overset{\perp}{Z}_0, \overset{\perp}{Z}_1, \dots$  on this space with absorption at  $x_{-n}, x_n$ . Use the argument that  $Z_k = |\overset{\perp}{Z}_k - x_0| + x_0$  is the reflected random walk we want, to see that for  $x_k \neq x_0$ ,

$$n^{(r)}(x_j; x_k) = \overset{\perp}{n}(x_j; x_k) + \overset{\perp}{n}(x_j; x_{-k}),$$

where  $\overset{\perp}{n}(x_j; x_k)$  refers to the  $\overset{\perp}{Z}$  random walk, and

$$n^{(r)}(x_j; x_0) = \overset{\perp}{n}(x_j; x_0).$$

Hence, for  $G_J(x, y)$  as defined in (16.46),

$$n^{(r)}(x_j; x_k) = \begin{cases} \frac{1}{\delta} G_J(x_j, x_k), & k \neq 0, \\ \frac{1}{2\delta} G_J(x_j, x_0), & k = 0. \end{cases}$$

For  $m(x_k, J_k)$ ,  $k \geq 1$ , use the expression  $\int_{J_k} G_{J_k}(x_k, y)m(dy)$ . Then

$$m(x_j, J) = \frac{1}{2\delta} G_J(x_j, x_0)m(x_0, J_0) + \int_{(b, c)} \left( \sum_{k=1}^{n-1} \frac{1}{\delta} G_J(x_j, x_k) G_{J_k}(x_k, y) \right) m(dy).$$

The integrand converges uniformly to  $G_J(x_j, y)$  as  $\delta \rightarrow 0$ . Hence for  $b$  not an exit boundary, the second term of this converges to

$$\int_{J - \{b\}} G_J(x_j, y)m(dy)$$

as we pass through successive refinements such that  $\delta \rightarrow 0$ . Therefore,  $m(x_0, J_0)/2\delta$  must converge to a limit, finite or infinite. Define  $m(\{b\})$  to be this limit. Then for  $x_j$  any point in the successive partitions,

$$m(x_j, J) = \int_J G_J(x_j, y)m(dy).$$

For  $m(\{b\}) < \infty$ , extend this to all  $x \in J$  by an argument similar to that in the proof of 16.36. For  $m(\{b\}) = \infty$ , show in the same way that  $m(x, J) = \infty$  for all  $x \in J$ .

Note that the behavior of the process is specified at all boundary points except regular boundary points by  $m(dy)$  on  $\mathcal{B}_1(\text{int } F)$ . Summarizing graphically, for  $b$  a right-hand endpoint,

Classification	Type
$\text{int}(F) \not\leftrightarrow b$	natural boundary
$\text{int}(F) \xleftarrow{r} b$	entrance boundary
$\text{int}(F) \xrightarrow{r} b$	exit boundary
$\text{int}(F) \rightleftarrows b$	regular boundary

The last statement ( $\rightleftarrows$  for regular boundary points) needs some explanation. Consider reflecting and absorbing Brownian motion on  $[b, \infty)$  as described in Section 3. Both of these processes have the same speed measure  $m(dy) = dy$  on  $(b, \infty)$ , and  $b$  is a regular boundary point for both of them. They differ in the assignment of  $m(\{b\})$ . For the absorbing process, obviously  $m(\{b\}) = \infty$ ; for the reflecting process  $m(\{b\}) < \infty$ . Hence, in terms of  $m$  on  $\text{int}(F)$  it is possible to go  $\text{int}(F) \rightarrow b$  and  $b \rightarrow \text{int}(F)$ . Of course, the latter is ruled out if  $m(\{b\}) = \infty$ , so  $\rightleftarrows$  should be understood above to mean "only in terms of  $m$  on  $\text{int}(F)$ ."

**Definition 16.49.** A regular boundary point  $b$  is called

absorbing	if $m(\{b\}) = \infty$ ,
slowly reflecting	if $0 < m(\{b\}) < \infty$ ,
instantaneously reflecting	if $m(\{b\}) = 0$ .

See Problem 11 for some interesting properties of a slowly reflecting boundary.

**Problems**

11. Show, by using the random walk approximation, that

$$(c - b)m(\{b\}) = E_b(l\{t; X(t) = b, t \leq \tau_c^*\}).$$

Conclude that if  $m(\{b\}) > 0$ , then  $X(t)$  spends a positive length of time at point  $b \Rightarrow p_t(\{b\} | b) > 0$  for some  $t$ . Show also that for almost all sample paths,  $\{t; X(t) = b\}$  contains no intervals with positive length and no isolated points.

12. For  $b$  an entrance boundary,  $J = (b, c)$ ,  $c \in \text{int}(F)$ , show that

$$m(b, J) = \int_J (c - y)m(dy),$$

where  $m(b, J) = \lim_{x \rightarrow b} m(x, J)$ .

13. For  $b$  a regular boundary point,  $J = [b, c)$ ,  $J \subset F$ , show that

$$m(b, J) = \int_J |y - c| m(dy).$$

14. For  $b$  an exit boundary, and any  $J = (a, b]$ ,  $x \in J$ , show that  $m(x, J) = \infty$  (see the proof of 16.47). Use 16.25 to conclude that  $P_b(\tau_x^* < \infty) = 0$ . Hence, deduce that  $p_t(\{b\} | b) = 1$ ,  $t \geq 0$ .

**8. CONSTRUCTION OF FELLER PROCESSES**

Assume that a Feller process is on its natural scale. Because  $X(t)$  has the same exit probabilities as Brownian motion  $X_0(t)$ , we should be able to construct a process with the same distribution as  $X(t)$  by expanding or contracting the time scale of the Brownian motion, depending on the current position of the particle. Suppose  $m(dx)$  is absolutely continuous with respect to Lebesgue measure,

$$m(dx) = V(x) dx,$$

where  $V(x)$  is continuous on  $F$ . For  $J$  small,

$$m(x, J) \simeq V(x)m^{(0)}(x, J),$$

where  $m^{(0)}(x, J)$  is the expected time for  $X_0(t)$  to exit from  $J$ . So if it takes Brownian motion time  $\Delta t$  to get out of  $J$ , it takes  $X(t)$  about  $V(x) \Delta t$  to get out of  $J$ . Look at the process  $X_0(T(t))$ , where  $T(t) = T(t, \omega)$  is for every  $\omega$  an increasing function of time. If we want this process to look like  $X(t)$ , then when  $t$  changes by  $V(x) \Delta t$ , we must see that  $T$  changes by the amount  $\Delta t$ . We get the differential equation

$$\frac{dT}{dt} = \frac{1}{V(x)}.$$

We are at the point  $x = X_0(T(t))$ , so

$$\frac{dT}{dt} = \frac{1}{V(X_0(T))} \quad \text{or} \quad V(X_0(T)) dT = dt.$$

Integrating, we get

$$\int_0^T V(X_0(\xi)) d\xi = t.$$

Hence,

**Definition 16.50.** Take  $X_0(t)$  to be Brownian motion on  $F$ , instantaneously reflecting at all finite endpoints. Denote

$$l(r) = \int_0^r V(X_0(\xi)) d\xi.$$

Define  $T(t)$  to be the solution of  $l(r) = t$ , that is,

$$t = \int_0^{T(t)} V(X_0(\xi)) d\xi.$$

**Remark.** Because  $m(J) > 0$  for all open neighborhoods  $J$ ,  $\{x; V(x) = 0\}$  can contain no open intervals. But almost no sample function of  $X_0(t)$  can have any interval of constancy. Hence  $l(\tau)$  is a strictly increasing continuous function of  $\tau$ , and so is  $T(t)$ . Further, note that for every  $t$ ,  $T(t)$  is a Markov time for  $\{X_0(t)\}$ .

**Theorem 16.51.**  $X(t) = X_0(T(t))$  is a Feller process on natural scale with speed measure

$$m(dx) = V(x) dx.$$

*Proof.* For any Markov time  $t^*$ ,  $X_0(t^*)$  is measurable  $\mathcal{F}(X_0(s), s \leq t^*)$ . Further, it is easy to show for any two Markov times  $t_1^* \leq t_2^*$ , that

$$\mathcal{F}(X_0(s), s \leq t_1^*) \subset \mathcal{F}(X_0(s), s \leq t_2^*).$$

This implies that  $X(\tau) = X_0(T(\tau))$  is measurable  $\mathcal{F}(X_0(s), s \leq T(t))$ , for any  $\tau \leq t$ . So

$$\mathcal{F}(X(s), s \leq t) \subset \mathcal{F}(X_0(s), s \leq T(t)).$$

Hence  $P_x(X(t + \tau) \in B \mid X(s), s \leq t)$  equals the expectation of

$$P_z(X(t + \tau) \in B \mid X_0(s), s \leq T(t)),$$

given  $\mathcal{F}(X(s), s \leq t)$ . To evaluate  $T(t + \tau)$ , write

$$t + \tau = \int_0^{T(t+\tau)} V(X_0(\xi)) d\xi = \int_0^\Delta V(X_0(\xi + T(t))) d\xi + t,$$

where  $\Delta = T(t + \tau) - T(t)$ . Thus  $\Delta$  is the solution of  $l_1(r) = \tau$ , where

$$l_1(r) = \int_0^r V(X_0(\xi + T(t))) d\xi.$$



Now  $X(t + \tau) = X_0(T(t + \tau)) = X_0(\Delta + T(t))$ . Because  $l_1(r)$  is the same function on the process  $X_0(\cdot + T(t))$  as  $l(r)$  is on  $X_0(\cdot)$ , the strong Markov property applies:

$$\begin{aligned} P_x(X_0(\Delta + T(t)) \in B \mid X_0(s), s \leq T(t)) \\ = P_{X_0(T(t))}(X_0(T(t)) \in B) = P_{X(t)}(X(\tau) \in B). \end{aligned}$$

The proof that  $X(t)$  is strong Markov is only sketched. Actually, what we will prove is that the transition probabilities for the process are stable. Let  $\varphi(x)$  be bounded and continuous on  $F$ . Denote by  $\tau_y^*$  the first passage time of  $X_0(t)$  to  $y$ . As  $y \rightarrow x$ ,  $\tau_y^* \rightarrow 0$  a.s.  $P_x$ . Hence, by path continuity, as  $y \rightarrow x$ ,

$$E_x \varphi(X(t + l(\tau_y^*))) \rightarrow E_x \varphi(X(t)).$$

$T(t + l(\tau_y^*))$  is defined as the solution of

$$t + l(\tau_y^*) = \int_0^t V(X_0(\xi)) d\xi.$$

Thus  $\Delta = T(t + l(\tau_y^*)) - \tau_y^*$  is the solution of

$$t = \int_0^\Delta V(X_0(\xi + \tau_y^*)) d\xi.$$

Use the strong Markov property of  $X_0(t)$  to compute

$$E_x \varphi(X(t + l(\tau_y^*))) = E_x \varphi(X_0(\Delta + \tau_y^*)) = E_y \varphi(X(t)).$$

Thus  $E_x \varphi(X(t))$  is continuous.

Denote by  $\tau_0^*(J)$  the first exit time of  $X_0(t)$  from  $J$ ,  $\tau^*(J)$  the exit time for  $X(t)$ . The relationship is

$$(16.52) \quad \tau_0^*(J) = T(\tau^*(J)) \quad \text{or} \quad \tau^*(J) = \int_0^{\tau_0^*(J)} V(X_0(\xi)) d\xi.$$

Taking expectations, we get

$$m(x, J) = E_x \int_0^{\tau_0^*(J)} V(X_0(\xi)) d\xi.$$

By Problem 10, the latter integral is

$$\int_J G_J(x, y) V(y) m_0(dy) = \int_J G_J(x, y) V(y) dy.$$

Thus, the process has the asserted speed measure.

To show  $X(t)$  is on its natural scale, take  $J = (a, b)$ ,

$$P_x(X(\tau^*(J)) = b) = P_x(X_0(T(\tau^*(J))) = b) = P_x(X_0(\tau_0^*(J)) = b).$$

If  $m(dx)$  is not absolutely continuous, the situation is much more difficult. The key difficulty lies in the definition of  $l(r)$ . So let us attempt to transform

the given definition of  $l(r)$  above into an expression not involving  $V(x)$ :  
Let  $L(t, J)$  be the time that  $X_0(\xi)$  spends in  $J$  up to time  $t$ ,

$$L(t, J) = l\{\xi; X_0(\xi) \in J, \xi \leq t\}.$$

Then  $l(r)$  can be written approximately as

$$l(r) = \sum_k V(x_k) L(t, J_k) = \sum_k \frac{L(t, J_k)}{\|J_k\|} m(J_k).$$

Suppose that

$$\lim_{J \downarrow \{y\}} \frac{L(t, J)}{\|J\|} = l^*(t, y)$$

exists for all  $y, t$ . Then assuming the limit exists in some nice way in  $y$ , we get the important alternative expression for  $l(r)$ ,

$$(16.53) \quad l(r) = \int l^*(r, y) m(dy).$$

That such a function  $l^*(t, x)$  exists having some essential properties is the consequence of a remarkable theorem due to Trotter [136].

**Theorem 16.54.** *Let  $X_0(t)$  be unrestricted Brownian motion on  $R^{(1)}$ . Almost every sample path has the property that there is a function  $l^*(t, y)$  continuous on  $\{t, y\}; t \geq 0, y \in R^{(1)}\}$  such that for all  $B \in \mathcal{B}_1$ ,*

$$l\{\xi; X_0(\xi) \in B, \xi \leq t\} = \int_B l^*(t, y) dy.$$

**Remarks.**  $l^*(t, y)$  is called local time for the process. It has to be non-decreasing in  $t$  for  $y$  fixed. Because of its continuity, the limited procedure leading up to (16.53) is justified. The proof of this theorem is too long and technical to be given here. See Ito and McKean [76, pp. 63 ff.] or Trotter [136].

Assume the validity of 16.54. Then it is easy to see that local time  $l_1^*(t, y)$  exists for the Brownian motion with reflecting boundaries. For example, if  $x = 0$  is a reflecting boundary, then for  $B \in \mathcal{B}_1([0, \infty))$ ,

$$l\{\xi; |X_0(\xi)| \in B, \xi \leq t\} = \int_B l_1^*(t, y) dy,$$

where  $l_1^*(t, y) = l^*(t, y) + l^*(t, -y)$ .

**Definition 16.55.** *Take  $X_0(t)$  to be Brownian motion on  $F$ , instantaneously reflecting at all closed endpoints, with local time  $l^*(t, y)$ . Let  $m(dx)$  be any measure on  $\mathcal{B}_1(F)$  such that  $0 < m(J) < \infty$  for all finite open intervals  $J$  with  $J \subset \text{int}(F)$ . Denote*

$$l(r) = \int l^*(r, y) m(dy),$$

and let  $T(t)$  be the solution of  $l(r) = t$ .

Since  $m(F)$  is not necessarily finite, then along some path there may be an  $r$  such that  $l(r) = \infty$ ; hence  $l(s) = \infty$ ,  $s \geq r$ . But if  $l(r) < \infty$ , then  $l(s)$  is continuous on  $0 \leq s \leq r$ . Furthermore, it is strictly increasing on this range. Otherwise there would be an  $s, t$ ,  $0 \leq t < s \leq r$  such that  $l^*(t, y) = l^*(s, y)$  for all  $y \in F$ . Integrate this latter equality over  $F$  with respect to  $dy$  to get the contradiction  $s = t$ . Thus,  $T(t)$  will always be well defined except along a path such that there is an  $r$  with  $l(r) < \infty$ ,  $l(r+) = \infty$  (by the monotone convergence theorem  $l(r)$  is always left-continuous). If this occurs, define  $T(t) \equiv r$  for all  $t \geq l(r)$ . With this added convention,  $T(t)$  is continuous and strictly increasing unless it becomes identically constant.

**Theorem 16.56.**  $X_0(T(t))$  is a Feller process on natural scale with speed measure  $m(dy)$ .

*Proof.* That  $X(t) = X_0(T(t))$  is on natural scale is again obvious. To compute the speed measure of  $X(t)$ , use (16.52),  $t_0^*(J) = T(t^*(J))$  or  $l(t_0^*(J)) = t^*(J)$ . Hence

$$m(x, J) = \int E_x l^*(t_0^*(J), y) m(dy).$$

The integrand does not depend on  $m(dy)$ . By the definition and continuity properties of local time the integral

$$\int_I E_x l^*(t_0^*(J), y) dy, \quad I \subset J$$

is the expected amount of time the Brownian particle starting from  $x$  spends in the interval  $I$  before it exits from  $J$ . We use Problem 10 to deduce that this expected time is also given by

$$\int_I G_J(x, y) dy.$$

The verification that  $E_x l^*(t_0^*(J), y)$  is continuous in  $y$  leads to its identification with  $G_J(x, y)$ , and proves the assertion regarding the speed measure of the process. Actually, the identity of Problem 10 was asserted to hold only for the interior of  $F$ , but the same proof goes through for  $J$  including a closed endpoint of  $F$  when the extended definition of  $G_J(x, y)$  is used.

The proof in 16.51 that  $X(t)$  is strong Markov is seen to be based on two facts:

- a)  $T(t)$  is a Markov time for every  $t \geq 0$ ;
- b)  $T(t + \tau) = \Delta + T(t)$ , where  $\Delta$  is the solution of  $l_1(r) = \tau$ , and  $l_1(r)$  is the integral  $l(r)$  based on the process  $X_0(\cdot + T(t))$ .

It is easy to show that (a) is true in the present situation. To verify (b), observe that  $l^*(t, y)$  is a function of  $y$  and the sample path for  $0 \leq \xi \leq t$ .

Because

$$\begin{aligned} I\{\xi; X_0(\xi) \in B, \xi \leq r + s\} \\ = I\{\xi; X_0(\xi) \in B, \xi \leq r\} + I\{\xi; X_0(\xi + r) \in B, 0 \leq \xi \leq s\}, \end{aligned}$$

it follows that

$$l^*(r + s, y) = l^*(r, y) + l^{**}(s, y),$$

where  $l^{**}(s, y)$  is the function  $l^*(s, y)$  evaluated along the sample path  $X_0(r + \xi)$ ,  $0 \leq \xi \leq s$ . Therefore

$$l(r + s) = l(r) + l_1(s),$$

where  $l_1(s)$  is  $l(s)$  evaluated along the path  $X_0(r + \xi)$ ,  $0 \leq \xi \leq s$ . The rest goes as before.

Many details in the above proof are left to the reader to fill in. An important one is the examination of the case in which  $l(r) = \infty$  for finite  $r$ . This corresponds to the behavior of  $X(t)$  at finite boundary points. In particular, if at a closed boundary point  $b$  the condition of accessibility (16.43) is violated, then it can be shown that the constructed process never reaches  $b$ . Evidently, for such measures  $m(dy)$ , the point  $b$  should be deleted from  $F$ . With this convention we leave to the reader the proof that the constructed process  $X(t)$  is regular on  $F$ .

**Problem 15.** Since  $l^*(t, 0)$  is nondecreasing, there is an associated measure  $l^*(dt, 0)$ . Show that  $l^*(dt, 0)$  is concentrated on the zeros of  $X_0(t)$ . That is, prove

$$\int_{\{\xi; X_0(\xi)=0, \xi \leq t\}} l^*(d\xi, 0) = l^*(t, 0).$$

(This problem illustrates the fact that  $l^*(t, 0)$  is a measure of the time the particle spends at zero.)

## 9. THE CHARACTERISTIC OPERATOR

The results of the last section show that corresponding to every speed measure  $m(dx)$  there is at least one Feller process on natural scale. In fact, there is only one. Roughly, the reasoning is that by breaking  $F$  down into smaller and smaller intervals, the distribution of time needed to get from one point to another becomes determined by the expected times to leave the small subintervals, hence by  $m(dx)$ . But to make this argument firm, an excursion is needed. This argument depends on what happens over small space intervals. The operator  $S$  which determines behavior over small time intervals can be computed by allowing a fixed time  $t$  to elapse, averaging over  $X(t)$ , and then taking  $t \rightarrow 0$ . Another approach is to fix the terminal space positions and average over the time it takes the particle to get to these

terminal space positions. Take  $x \in J$ ,  $J$  open, and define

$$(16.57) \quad (U_J f)(x) = \frac{E_x f(X(t^*(J))) - f(x)}{E_x t^*(J)}.$$

Let  $J \downarrow \{x\}$ , if  $\lim (U_J f)(x)$  exists, then see whether, under some reasonable conditions, the limit will equal  $(Sf)(x)$ .

**Definition 16.58.** Let  $x \in F$ . For any neighborhood  $J$  of  $x$  open relative to  $F$ , suppose a function  $\varphi(J)$  is defined. Say that

$$\lim_{J \downarrow \{x\}} \varphi(J) = \alpha$$

if for every system  $J_n$  of such neighborhoods,  $J_n \downarrow \{x\}$ ,  $\lim_n \varphi(J_n) = \alpha$ .

**Theorem 16.59.** Let  $f \in \mathcal{D}(S, \mathcal{C})$ , where  $\mathcal{C}$  consists of all bounded continuous functions on  $F$ . Then

$$\lim_{J \downarrow \{x\}} (U_J f)(x) = (Sf)(x)$$

for all  $x \in F$ .

*Proof.* The proof is based on an identity due to Dynkin.

**Lemma 16.60.** For any Markov time  $t^*$  such that  $E_x t^* < \infty$ , for  $f \in \mathcal{D}(S, \mathcal{C})$  and  $g = Sf$ ,

$$E_x f(X(t^*)) - f(x) = E_x \int_0^{t^*} g(X(t)) dt.$$

*Proof.* For any bounded measurable  $h(x)$  consider  $f(x) = (R_\lambda h)(x)$  and write

$$\begin{aligned} (16.61) \quad E_x e^{-\lambda t^*} f(X(t^*)) &= E_x \int_0^\infty e^{-\lambda(t+t^*)} h(X(t+t^*)) dt \\ &= E_x \int_{t^*}^\infty e^{-\lambda t} h(X(t)) dt \\ &= f(x) - E_x \int_0^{t^*} e^{-\lambda t} h(X(t)) dt. \end{aligned}$$

For  $f$  in  $\mathcal{D}(S, \mathcal{C})$ , take  $h = (\lambda - S)f$ . Then by 15.51(2),  $f = R_\lambda h$  and (16.61) becomes

$$E_x e^{-\lambda t^*} f(X(t^*)) - f(x) = E_x \int_0^{t^*} e^{-\lambda t} g(X(t)) dt - \lambda E_x \int_0^{t^*} e^{-\lambda t} f(X(t)) dt,$$

where the last integral exists for all  $\lambda \geq 0$  by the boundedness of  $f(x)$  and  $E_x t^* < \infty$ . Taking  $\lambda \rightarrow 0$  now proves the lemma.

To finish 16.59, if  $x$  is an absorbing or exit endpoint, then both expressions are zero at  $x$ . For any other point  $x$ , there are neighborhoods  $J$

of  $x$ , open relative to  $F$ , such that  $E_x \tau^*(J) < \infty$ . Now  $g$  is continuous at  $x$ , so take  $J$  sufficiently small so that  $|g(X(t)) - g(x)| \leq \epsilon$ ,  $t \leq \tau^*(J)$ . Then by the lemma,

$$E_x f(X(\tau^*(J))) - f(x) = (g(x) \pm \epsilon) E_x \tau^*(J).$$

**Definition 16.62.** Define  $(Uf)(x)$  for any measurable function  $f$  as

$$\lim_{J \downarrow \{x\}} (U_J f)(x)$$

wherever this limit exists. Denote  $f \in \mathcal{D}(U, I)$  if the limit exists for all  $x \in I$ .

**Corollary 16.63.** If  $f \in \mathcal{D}(S, \mathbb{C})$ , then  $f \in \mathcal{D}(U, F)$ .

There are a number of good reasons why  $U$  is usually easier to work with than  $S$ . For our purposes an important difference is that  $U$  is expressible in terms of the scale and speed measure. For any Feller process we can show that  $(Uf)(x)$  is very nearly a second-order differential operator. There is a simple expression for this when the process is on natural scale.

**Theorem 16.64.** Let  $X(t)$  be on its natural scale,  $f \in \mathcal{D}(S, \mathbb{C})$ . Then

$$(Uf)(x) = \frac{1}{2} \frac{d}{dm} \frac{d}{dx} f(x)$$

in the following sense for  $x \in \text{int}(F)$ ,

- i)  $f'(x)$  exists except perhaps on the countable set of points  $\{x; m(\{x\}) > 0\}$ .
- ii) For  $J = (x_1, x_2)$  finite, such that  $f'(x_1), f'(x_2)$  exist,

$$f'(x_2) - f'(x_1) = 2 \int_J g(x) m(dx),$$

where  $g(x) = (Sf)(x)$ .

*Proof.* Use 16.60 and Problem 10 to get

$$E_x f(X(\tau^*(J))) - f(x) = \int_J G_J(x, y) g(y) m(dy).$$

Take  $J = (x - h_1, x + h_2)$  and use the appropriate values for  $p^+(x, J)$  to get the following equations.

$$\frac{f(x + h_2) - f(x)}{h_2} - \frac{f(x) - f(x - h_1)}{h_1} = \frac{h_2 + h_1}{h_1 h_2} \int G_J(x, y) g(y) m(dy).$$

$$\frac{h_2 + h_1}{2h_1 h_2} G_J(x, y) = \begin{cases} \frac{y - (x - h_1)}{h_1}, & x - h_1 \leq y \leq x, \\ \frac{(x + h_2) - y}{h_2}, & x < y \leq x + h_2. \end{cases}$$

Taking  $h_1, h_2 \downarrow 0$  separately we can show that both right- and left-hand derivatives exist. When  $h_1, h_2$  both go to zero, the integrand converges to zero everywhere except at  $y = x$ . Use bounded convergence to establish (i).

By substituting  $x + h_2 = b, x - h_1 = a$ , form the identity,

$$(x - a)f(b) + (b - x)f(a) - f(x)(b - a) = (b - a) \int G_J(x, y)g(y)m(dy).$$

Substituting in  $x + h$  for  $x$  and subtracting we get

$$\begin{aligned} f(b) - f(a) - (b - a) \frac{f(x + h) - f(x)}{h} \\ = (b - a) \int \frac{G_J(x + h, y) - G_J(x, y)}{h} g(y)m(dy). \end{aligned}$$

The integrand is bounded for all  $h$ , and

$$\lim_{h \downarrow 0} (b - a) \frac{G_J(x + h, y) - G_J(x, y)}{h} = \begin{cases} -2(y - a), & y < x, \\ 2(b - y), & y > x. \end{cases}$$

Therefore  $g(x)m(\{x\}) = 0$  implies that

$$f(b) - f(a) - (b - a)f'(x) = -2 \int_a^x (y - a)g \, dm + 2 \int_x^b (b - y)g \, dm.$$

Take  $a < x_1 < x_2 < b$ , such that

$$g(x_1)m(\{x_1\}) = g(x_2)m(\{x_2\}) = 0.$$

Use both  $x_1$  and  $x_2$  in the equation above and subtract

$$\begin{aligned} (b - a)(f'(x_2) - f'(x_1)) &= 2 \int_{x_1}^{x_2} (y - a)g \, dm + 2 \int_{x_2}^{x_1} (b - y)g \, dm \\ &= 2(b - a) \int_{x_1}^{x_2} g \, dm. \end{aligned} \quad \text{Q.E.D.}$$

There is an interesting consequence of this theorem in an important special case:

**Corollary 16.65.** Let  $m(dx) = V(x) dx$  on  $\mathcal{B}(\text{int } F)$  where  $V(x)$  is continuous on  $\text{int}(F)$ . Then  $f \in \mathcal{D}(S, \mathbb{C})$  implies that  $f''(x)$  exists and is continuous on  $\text{int}(F)$ , and for  $x \in \text{int}(F)$ ,

$$(Uf)(x) = \frac{1}{2V(x)} \frac{d^2 f(x)}{dx^2}.$$

**Problems**

16. If  $b$  is a closed left endpoint,  $f \in \mathcal{D}(S, \mathbb{C})$ , show that

$$b \text{ reflecting} \Rightarrow f'_+(b) = m(\{b\})(Sf)(b),$$

where  $f'_+(b)$  is the one-sided derivative

$$\lim_{h \downarrow 0} \frac{f(b+h) - f(b)}{h}, \quad b+h \in F.$$

Show also that

$$b \text{ absorbing or exit} \Rightarrow (Sf)(b) = 0.$$

17. If  $X(t)$  is not on its natural scale, use the definition

$$g(x) = \left( \frac{df}{du} \right)(x) \Leftrightarrow f(x_2) - f(x_1) = \int_{(x_1, x_2)} g(x) du(x).$$

Show that for  $f \in \mathcal{D}(S, \mathbb{C})$ ,  $x \in \text{int}(F)$ ,

$$(Uf)(x) = \frac{1}{2} \left( \frac{d}{dm} \frac{df}{du} \right)(x).$$

**10. UNIQUENESS**

$U$  is given by the scale and speed. But the scale and speed can also be recovered from  $U$ .

**Proposition 16.66.**  $p^+(x, J)$  satisfies

$$(Up^+)(x) = 0, \quad x \in J;$$

and  $m(x, J)$  satisfies

$$(Um)(x) = -1, \quad x \in J,$$

for  $J$  open and finite,  $J \subset F$ .

*Proof.* Let  $I \subset J$ ,  $I = (x_1, x_2)$ . Then for  $x \in I$ ,

$$E_x p^+(X(t^*(I)), J) = p^+(x_2, J)p^+(x, I) + p^+(x_1, J)p^-(x, I) = p^+(x, J),$$

$$E_x m(X(t^*(I), J)) = m(x_2, J)p^+(x, I) + m(x_1, J)p^-(x, I) = m(x, J) - m(x, I).$$

For  $b$  a closed reflecting endpoint and  $J = [b, c)$ , even simpler identities hold and 16.66 is again true for  $J$  of this form.

To complete the recovery of  $m, p^+$  from  $U$ , one needs to know that the solutions of the equations in 16.66 are unique, subject to the conditions  $m, p^+$  continuous on  $J = (a, b)$ , and  $p^+(b-, J) = 1$ ,  $p^+(a+, J) = 0$ ,  $m(a+, J) = 0$ ,  $m(b-, J) = 0$ .



**Proposition 16.67.** Let  $J = (a, b)$ . If  $f(x)$  is continuous on  $J$ , then

$$(Uf)(x) = 0, \quad x \in J, \quad f(a) = f(b) = 0 \Rightarrow f(x) \equiv 0, \quad x \in J.$$

*Proof.* This is based on the following minimum principle for  $U$ :

**Lemma 16.68.** If  $h(x)$  is a continuous function in some neighborhood of  $x_0 \in F$ , if  $h \in \mathcal{D}(U, \{x_0\})$ , and if  $h(x)$  has a local minimum at  $x_0$ , then

$$(Uh)(x_0) \geq 0.$$

This is obvious from the expression for  $U$ . Now suppose there is a function  $\varphi(x)$  continuous on  $J$  such that  $\varphi > 0$  on  $J$  and  $U\varphi = \lambda\varphi$  on  $J$ ,  $\lambda > 0$ . Suppose that  $f(x)$  has a minimum in  $J$ , then for  $\epsilon > 0$  sufficiently small,  $f - \epsilon\varphi$  has a minimum in  $J$ , but

$$U(f - \epsilon\varphi) = -\lambda\epsilon\varphi < 0 \quad \text{on } J.$$

By the minimum principle,  $f$  cannot have a minimum. Similarly it cannot have a maximum, thus  $f \equiv 0$ . All that is left to do is get  $\varphi$ . This we will do shortly in Theorem 16.69 below.

This discussion does not establish the fact that  $U$  determines  $m(\{b\})$  at a closed, reflecting endpoint  $b$ , because if  $J = [b, c)$ , then  $m(b, J) \neq 0$ . In this case suppose there are two solutions,  $f_1, f_2$  of  $Uf = -1$ ,  $x \in J$ , such that both are continuous on  $J$ , and  $f_1(c) = f_2(c) = 0$ ,  $f_1(b) = \alpha$ ,  $f_2(b) = \beta > \alpha$ . Form

$$f(x) = f_2(x) - \frac{\beta}{\alpha} f_1(x) \quad \text{so} \quad Uf = \frac{\beta}{\alpha} - 1 > 0.$$

The minimum principle implies that  $f(x) \leq 0$  on  $J$ . But  $(Uf)(b) > 0$  implies there is a number  $d$ ,  $b < d < c$ , such that  $f(d) - f(b) > 0$ . Hence  $\beta = \alpha$ ,  $f_1(x) = f_2(x)$  on  $J$ .

**Theorem 16.69.**  $U\varphi = \lambda\varphi$ ,  $x \in \text{int}(F)$ ,  $\lambda > 0$ , has two continuous solutions  $\varphi_+(x)$ ,  $\varphi_-(x)$  such that

- i)  $\varphi_+(x) > 0$ ,  $\varphi_-(x) > 0$ ,  $x \in \text{int}(F)$ ,
- ii)  $\varphi_+(x)$  is strictly increasing,  $\varphi_-(x)$  is strictly decreasing,
- iii) at closed endpoints  $b$ ,  $\varphi_+$  and  $\varphi_-$  have finite limits as  $x \rightarrow b$ , at open right (left) endpoints  $\varphi_+(\varphi_-) \rightarrow \infty$ ,
- iv) any other continuous solution of  $U\varphi = \lambda\varphi$ ,  $x \in \text{int}(F)$  is a linear combination of  $\varphi_+$ ,  $\varphi_-$ .

*Proof.* The idea is simple in the case  $F = [a, b]$ . Here we could take

$$\varphi_+(x) = E_x e^{-\lambda t^*}, \quad \varphi_-(x) = E_x e^{-\lambda t^*}$$

and show that these two satisfy the theorem. In the general case, we have

to use extensions of these two functions. Denote for  $x < z$ ,

$$\begin{aligned}\theta_+(x, z) &= E_x e^{-\lambda t^*}, \\ \theta_-(z, x) &= E_z e^{-\lambda t_x^*}.\end{aligned}$$

For  $x < y < z$ , the time from  $x$  to  $z$  consists of the time from  $x$  to  $y$  plus the time from  $y$  to  $z$ . Use the strong Markov property to get

$$(16.70) \quad \theta_+(x, z) = \theta_+(x, y)\theta_+(y, z), \quad \theta_-(z, x) = \theta_-(z, y)\theta_-(y, x).$$

Now, in general,  $t_x^*$  and  $t_z^*$  are extended random variables and therefore not stopping times. But, by truncating them and then passing to the limit, identities such as (16.70) can be shown to hold. Pick  $z_0$  in  $F$  to be the closed right endpoint if there is one, otherwise arbitrary. Define, for  $x \leq z_0$ ,

$$\varphi_+(x) = E_x e^{-\lambda t_{z_0}^*}.$$

Now take  $z_1 > z_0$ , and for  $x \leq z_1$  define

$$\varphi_+(x) = E_x e^{-\lambda t_{z_1}^*} / E_{z_0} e^{-\lambda t_{z_1}^*}.$$

Use (16.70) to check that the  $\varphi_+$  as defined for  $x \leq z_1$  is an extension of  $\varphi_+$  defined for  $x \leq z_0$ .

Continuing this way we define  $\varphi_+(x)$  on  $F$ . Note that  $\varphi_+ > 0$  on  $\text{int}(F)$ , and is strictly increasing. Now define  $\varphi_-$  in an analogous way. As  $w \uparrow y$  or  $w \downarrow y$ ,  $t_w^* \rightarrow t_y^*$  a.s.  $P_x$  on the set  $\{t_y^* < \infty\}$ . Use this fact applied to (16.70) to conclude that  $\varphi_+$  and  $\varphi_-$  are continuous. If the right-hand endpoint  $b$  is open, then  $\lim E_x \exp[-\lambda t_z^*] = 0$  as  $z \rightarrow b$ , otherwise  $b$  is accessible. For  $z > z_0$ , the definition of  $\varphi_+$  gives

$$\varphi_+(z) = 1/E_{z_0} e^{-\lambda t_z^*},$$

so  $\varphi_+(z) \rightarrow +\infty$  as  $z \rightarrow b$ .

For  $y$  arbitrary in  $F$ , let  $\varphi(x) = E_x e^{-\lambda t_y^*}$ , then assert that

**Proposition 16.71.**  $\varphi \in \mathcal{D}(U, F - \{y\})$ , and for  $x \neq y$ ,

$$U\varphi = \lambda\varphi.$$

*Proof.* Take  $h(x) \in \mathcal{C}$  such that  $h(x) = 0$  for all  $x \leq y$ . Then  $f = R_\lambda h$  is in  $\mathcal{D}(S, \mathcal{C})$ . By truncating, it is easy to show that the identity (16.61) holds for the extended stopping variable  $t_y^*$ . For  $x \leq y$ , (16.61) now becomes

$$E_x e^{-\lambda t_y^*} f(X(t_y^*)) = f(x)$$

or

$$f(y)\varphi(x) = f(x), \quad x \leq y.$$

By 15.51,  $(\lambda - S)f = h$ , so

$$(Uf)(x) = \lambda f(x), \quad x \leq y.$$

But for  $x < y$ ,

$$(Uf)(x) = f(y)(U\varphi)(x),$$

which leads to  $U\varphi = \lambda\varphi$ . An obviously similar argument does the trick for  $x > y$ .

Therefore  $\varphi_+$ ,  $\varphi_-$  are solutions of  $U\varphi = \lambda\varphi$  in  $\text{int}(F)$ . Let  $\varphi$  be any solution of  $U\varphi = \lambda\varphi$  in  $\text{int}(F)$ . For  $x_1 < x_2$  determine constants  $c_1, c_2$  so that

$$\varphi(x_2) = c_1\varphi_+(x_2) + c_2\varphi_-(x_2),$$

$$\varphi(x_1) = c_1\varphi_+(x_1) + c_2\varphi_-(x_1).$$

This can always be done if  $\varphi_+(x_2)\varphi_-(x_1) - \varphi_-(x_2)\varphi_+(x_1) \neq 0$ . The function  $D(x) = \varphi_+(x)\varphi_-(x_1) - \varphi_-(x)\varphi_+(x_1)$  is strictly increasing in  $x$ , so  $D(x_1) = 0 \Rightarrow D(x_2) > 0$ . The function  $\tilde{\varphi}(x) = \varphi(x) - c_1\varphi_+(x) - c_2\varphi_-(x)$  satisfies  $U\tilde{\varphi} = \lambda\tilde{\varphi}$ , and  $\tilde{\varphi}(x_1) = \tilde{\varphi}(x_2) = 0$ . The minimum principle implies  $\tilde{\varphi}(x) = 0$ ,  $x_1 \leq x \leq x_2$ . Thus

$$\varphi(x) = c_1\varphi_+(x) + c_2\varphi_-(x), \quad x_1 \leq x \leq x_2.$$

This must hold over  $\text{int}(F)$ . For if  $c'_1, c'_2$  are constants determined by a larger interval  $(x'_1, x'_2)$ , then

$$(c'_1 - c_1)\varphi_+(x) + (c'_2 - c_2)\varphi_-(x) = 0, \quad x_1 \leq x \leq x_2.$$

This is impossible unless  $\varphi_+, \varphi_-$  are constant over  $(x_1, x_2)$ . Hence  $c'_1 = c_1$ ,  $c'_2 = c_2$ .

We can now prove the uniqueness.

**Theorem 16.72.** For  $f(x)$  bounded and continuous on  $F$ ,  $g(x) = (R_\lambda f)(x)$ ,  $\lambda > 0$ , is in  $\mathcal{D}(U, F)$  and is the unique bounded continuous solution of

$$(Ug)(x) = \lambda g(x) + f(x), \quad \text{all } x.$$

*Proof.* That  $g(x) \in \mathcal{D}(U, F)$  follows from 16.63, and 15.51. Further,

$$Sg = \lambda g + f \Rightarrow Ug = \lambda g + f.$$

Suppose that there were two bounded continuous solutions  $g_1, g_2$  on  $F$ . Then  $\varphi(x) = g_1(x) - g_2(x)$  satisfies

$$U\varphi = \lambda\varphi, \quad x \in F.$$

By 16.69,

$$\varphi(x) = c_1\varphi_+(x) + c_2\varphi_-(x), \quad x \in \text{int}(F).$$

If  $F$  has two open endpoints,  $\varphi$  cannot be bounded unless  $c_1 = c_2 = 0$ .

Therefore, take  $F$  to have at least one closed endpoint  $b$ , say to the left. Assume that  $\varphi(b) \geq 0$ ; otherwise, use the solution  $-\varphi(x)$ . If  $\varphi(b) > 0$ , then

$$(U\varphi)(b) > 0 \Rightarrow \varphi(b+h) > \varphi(b)$$

for all  $h$  sufficiently small. Then  $\varphi(x)$  can never decrease anywhere, because if it did, it would have a positive maximum in  $\text{int}(F)$ , contradicting the minimum principle. If  $F = [b, c)$ , then  $\varphi(x) = c_2\varphi_-(x)$ . Since  $\varphi_-(b) = 1$ , the case  $\varphi(b) = 0$  leads to  $\varphi(x) \equiv 0$ . But  $\varphi(b) > 0$  is impossible because  $\varphi_-(x)$  is decreasing. Finally, look at  $F = [b, c]$ . If  $\varphi(c) > 0$  or if  $\varphi(c) < 0$ , then an argument similar to the above establishes that  $\varphi(x)$  must be decreasing on  $F$  or increasing on  $F$ , respectively. In either case,  $\varphi(b) \geq 0$  is impossible. The only case not ruled out is  $\varphi(b) = \varphi(c) = 0$ . Here the minimum principle gives  $\varphi(x) \equiv 0$  and the theorem follows.

**Corollary 16.73.** *There is exactly one Feller process having a given scale function and speed measure.*

*Proof.* It is necessary only to show that the transition probabilities  $\{p_t(B|x)\}$  are uniquely determined by  $u(x)$ ,  $m(dx)$ . This will be true if  $E_x f(X(t))$  is uniquely determined for all bounded continuous  $f$  on  $F$ . The argument used to prove uniqueness in 15.51 applies here to show that  $E_x f(X(t))$  is completely determined by the values of  $(R_\lambda f)(x)$ ,  $\lambda > 0$ . But  $g(x, \lambda) = (R_\lambda f)(x)$  is the unique bounded continuous solution of  $Ug = \lambda g + f$ ,  $x \in F$ .

## 11. $\varphi_+(x)$ AND $(\varphi_-x)$

These functions, (which depend also on  $\lambda$ ) have a central role in further analytic developments of the theory of Feller processes.

For example, let  $J$  be any finite interval, open in  $F$ , with endpoints  $x_1, x_2$ . The first passage time distributions from  $J$  can be specified by the two functions

$$\theta^+(\lambda, x) = \int_{A_+} e^{-\lambda t^*(J)} dP_x, \quad \theta^-(\lambda, x) = \int_{A_-} e^{-\lambda t^*(J)} dP_x,$$

where  $A_+ = \{X(t^*(J)) = x_2\}$ ,  $A_-$  is the other exit set.

**Theorem 16.74** (Darling and Siebert)

$$\begin{aligned} \theta^+(\lambda, x) &= \frac{\varphi_+(x_1)\varphi_-(x) - \varphi_+(x)\varphi_-(x_1)}{\varphi_+(x_1)\varphi_-(x_2) - \varphi_+(x_2)\varphi_-(x_1)}, \\ \theta^-(\lambda, x) &= \frac{\varphi_-(x_2)\varphi_+(x) - \varphi_-(x)\varphi_+(x_2)}{\varphi_+(x_1)\varphi_-(x_2) - \varphi_+(x_2)\varphi_-(x_1)}. \end{aligned}$$

**Remark.** Since these expressions are invariant under linear transformations, we could just as well take instead of  $\varphi_+, \varphi_-$  any two linearly independent solutions  $g_+, g_-$  of  $Ug = \lambda g$  such that  $g_+ \uparrow, g_- \downarrow$ .

*Proof.* Use the strong Markov property to write

$$E_x e^{-\lambda t_{x_2}^*} = \int_{A_+} e^{-\lambda t^*(J)} dP_x + (E_{x_1} e^{-\lambda t_{x_2}^*}) \int_{A_-} e^{-\lambda t^*(J)} dP_x,$$

$$E_x e^{-\lambda t_{x_1}^*} = (E_{x_2} e^{-\lambda t_{x_1}^*}) \int_A e^{-\lambda t^*(J)} dP_x + \int_{A_-} e^{-\lambda t^*(J)} dP_x.$$

Let  $h^+(x) = E_x e^{-\lambda t_{x_2}^*}$ ,  $h^-(x) = E_x e^{-\lambda t_{x_1}^*}$ , and solve the above equations to get

$$\theta^+(\lambda, x) = \frac{h^+(x) - h^+(x_1)h^-(x)}{1 - h^-(x_2)h^+(x_1)}.$$

Since  $h^-(x_1) = 1$ ,  $h^+(x_2) = 1$ , this has the form of the theorem. By construction  $h^-(x)$ ,  $h^+(x)$  are constant multiples of  $\varphi_-(x)$ ,  $\varphi_+(x)$ .

More important, in determining the existence of densities of the transition probabilities  $p_t(B|x)$  and their differentiability is the following sequence of statements:

**Theorem 16.75**

- i)  $\frac{d}{dx} \varphi_+(x)$ ,  $\frac{d}{dx} \varphi_-(x)$  both exist except possibly at a countable number of points.
- ii)  $\mathfrak{F} = \varphi_- \frac{d\varphi_-}{dx} - \varphi_+ \frac{d\varphi_+}{dx}$  is constant except possibly where  $\frac{d}{dx} \varphi_+$  or  $\frac{d}{dx} \varphi_-$  do not exist.
- iii)  $R_\lambda(dy|x) \ll m(dy)$  on  $\mathcal{B}_1(F)$ , and putting  $r_\lambda(y, x) = \frac{dR_\lambda}{dm}$ , we have

$$r_\lambda(y, x) = \begin{cases} \frac{1}{\mathfrak{F}} \varphi_+(y) \varphi_-(x), & y \leq x, \\ \frac{1}{\mathfrak{F}} \varphi_-(y) \varphi_+(x), & y \geq x, \end{cases}$$

or equivalently,

$$(R_\lambda f)(x) = \int f(y) r_\lambda(y, x) m(dy)$$

for  $r_\lambda(y, x)$  as above.

The proof, which I will not give (see Ito and McKean, pp. 149 ff.), goes like this: Show that  $U\varphi_+$ ,  $U\varphi_-$  have the differential form given in 16.64. Then statement (ii) comes from the fact that

$$\int_{(x_1, x_2)} (\varphi_- U\varphi_+ - \varphi_+ U\varphi_-) dm = 0.$$

Statement (iii) comes from showing directly that the function

$$g(x) = \int f(y) r_\lambda(y, x) m(dy)$$

satisfies  $Ug = \lambda g + f$  for all continuous bounded  $f$ .

The statement 16.75(iii) together with some eigenvalue expansions finally shows that  $p_t(dy | x) \ll m(dy)$  and that  $p_t(y | x)$ , the density with respect to  $m(dy)$  exists and is a symmetric function of  $y, x$ . Also, the same development shows that  $p_t(y | x) \in \mathcal{D}(S, \mathbb{C})$  and that

$$\frac{\partial}{\partial t} p_t(y | x) = (Sp_t(y | \cdot))(x).$$

Unfortunately, the proofs of these results require considerably more analytic work. It is disturbing that such basic things as existence of densities and the proofs that  $p_t(y | x)$  are sufficiently differentiable to satisfy the backwards equations lie so deep.

**Problem 18.** If regular Feller process has a stationary initial distribution  $\pi(dx)$  on  $F$ , then show that  $\pi(dx)$  must be a constant multiple of the speed measure  $m(dx)$ . [Use 16.75(iii).]

## 12. DIFFUSIONS

There is some disagreement over what to call a diffusion. We more or less follow Dynkin [44].

**Definition 16.76.** A diffusion is a Feller process on  $F$  such that there exist functions  $\sigma^2(x)$ ,  $\mu(x)$  defined and continuous on  $\text{int}(F)$ ,  $\sigma^2(x) > 0$  on  $\text{int}(F)$ , with

- i)  $\frac{1}{t} P_x(|X(t) - x| \geq \epsilon) \xrightarrow{bp} 0$ ,
- ii)  $\frac{1}{t} \int_{|X(t)-x| < \epsilon} (X(t) - x) dP_x \xrightarrow{bp} \mu(x)$ ,
- iii)  $\frac{1}{t} \int_{|X(t)-x| < \epsilon} (X(t) - x)^2 dP_x \xrightarrow{bp} \sigma^2(x)$ ,

where the convergence ( $\xrightarrow{bp}$ ) is bounded pointwise on all finite intervals  $J$ ,  $J \subset \text{int}(F)$ .

So a diffusion is pretty much what we started in with—a process that is locally Brownian. Note

**Proposition 16.77.** *Let  $f(x)$  have a continuous second derivative on  $\text{int}(F)$ . For a diffusion,  $f \in \mathcal{D}(U, \text{int}(F))$  and*

$$(Uf)(x) = \frac{1}{2}\sigma^2(x)\frac{d^2f}{dx^2} + \mu(x)\frac{df}{dx}.$$

*Proof.* Define  $f(x) = \tilde{f}(x)$  in some neighborhood  $J$  of  $x$  such that  $\tilde{f}(x)$  is bounded and has continuous second derivative on  $\text{int}(F)$  and vanishes outside a compact interval  $I$  contained in the interior of  $F$ . On  $I^c$  prove that  $(S_t \tilde{f})(x) \xrightarrow{b.p.} 0$ . Apply a Taylor expansion now to conclude that  $\tilde{f} \in \mathcal{D}(S, \mathbb{C})$ , and that  $(S\tilde{f})(x)$  is given by the right-hand side above.

The scale is given in terms of  $\mu, \sigma$  by

**Proposition 16.78.** *For a diffusion, the scale function  $u(x)$  is the unique (up to a linear transformation) solution on  $\text{int}(F)$  of*

$$(16.79) \quad \frac{1}{2}\sigma^2(x)\frac{d^2u(x)}{dx^2} + \mu(x)\frac{du(x)}{dx} = 0.$$

*Proof.* Equation (16.79) has the solution

$$(16.80) \quad \frac{du_0(x)}{dx} = \exp \left[ - \int_{x_0}^x \frac{2\mu(z)}{\sigma^2(z)} dz \right]$$

for an arbitrary  $x_0$ . This  $u_0(x)$  has continuous second derivative on  $\text{int}(F)$ . Thus by 16.77  $(Uu_0)(x) = 0$ ,  $x \in \text{int}(F)$ , and  $u_0(x)$  is a scale function.

Now that the scale is determined, transform to natural scale, getting the process

$$\tilde{X}(t) = u(X(t)).$$

**Proposition 16.81.**  *$\tilde{X}(t)$  is a diffusion with zero drift and*

$$\tilde{\sigma}^2(y) = \sigma^2(x)[u'(x)]^2, \quad y = u(x).$$

*Proof.* Proposition 16.81 follows from

**Proposition 16.82.** *Let  $X(t)$  be a diffusion on  $F$ ,  $w(x)$  a function continuous on  $F$  such that  $|w'(x)| \neq 0$  on  $\text{int}(F)$ ,  $w''(x)$  continuous on  $\text{int}(F)$ . Then  $\tilde{X}(t) = w(X(t))$  is a diffusion. If  $\tilde{x} = w(x)$ ,*

$$\tilde{\mu}(\tilde{x}) = \frac{1}{2}\sigma^2(x)w''(x) + \mu(x)w'(x),$$

$$\tilde{\sigma}^2(\tilde{x}) = \sigma^2(x)[w'(x)]^2.$$

*Proof.* The transformations of the drift and variance here come from

$$\begin{aligned}\frac{E_x w(X(t)) - w(x)}{t} &\rightarrow Sw(x), \\ \frac{E_x (w(X(t)) - w(x))^2}{t} &= \frac{E_x (w^2(X(t)) - w^2(x))}{t} - 2w(x) \cdot \frac{E_x w(X(t)) - w(x)}{t} \\ &\rightarrow Sw^2 - 2w \cdot Sw.\end{aligned}$$

What needs to be shown is that for the revised process,  $\xrightarrow{bp} \tilde{\mu}(\tilde{x})$ ,  $\xrightarrow{bp} \tilde{\sigma}^2(\tilde{x})$  takes place on finite intervals in the interior of  $\tilde{F}$ . This is a straightforward verification, and I omit it.

By this time it has become fairly apparent from various bits of evidence that the following should hold:

**Theorem 16.83.** *For a diffusion with zero drift the speed measure is given on  $\mathcal{B}_1(\text{int } F)$  by*

$$m(dx) = \frac{1}{\sigma^2(x)} dx.$$

*Proof.* Take  $f$  to be zero off a compact interval  $I \subset \text{int}(F)$ , with a continuous second derivative. There are two expressions, given by 16.77 and 16.64, for  $(Uf)(x)$ . Equating these,

$$\sigma^2(x) \frac{d^2 f(x)}{dx^2} = \left( \frac{d}{dm} \frac{df}{dx} \right)(x)$$

or,

$$\begin{aligned}f'(x_2) - f'(x_1) &= \int_{(x_1, x_2)} \sigma^2(y) f''(y) m(dy) \\ &= \int_{(x_1, x_2)} f''(y) dy\end{aligned}$$

which implies the theorem.

These results show that  $\mu(x)$ ,  $\sigma^2(x)$  determine the scale function and the speed measure completely on  $\text{int}(F)$ . Hence, specifying  $m$  on any and all regular boundary points completely specifies the process. The basic uniqueness result 16.73 guarantees at most one Feller process with the given scale and speed. Section 8 gives a construction for the associated Feller process. What remains is to show that the process constructed is a diffusion.

**Theorem 16.84.** *Let  $m(dx) = V(x) dx$  on  $\text{int}(F)$ ,  $V(x)$  continuous and positive on  $\text{int}(F)$ . The process on natural scale with this speed measure is a diffusion with  $\mu(x) = 0$ ,  $\sigma^2(x) = 1/V(x)$ .*



*Proof.* Use the representation  $X(t) = X_0(T(t))$ , where  $T(t)$  is the random time change of Section 8. The first step in the proof is: Let the interval  $J_\epsilon = (x - \epsilon, x + \epsilon)$ , take  $J \subset \text{int}(F)$  finite such that for any  $x \in J$ ,  $J_\epsilon \subset I \subset \text{int}(F)$ . Then

$$(i^\circ) \quad (1/t)P_x(\tau^*(J_\epsilon) \leq t) \xrightarrow{bp} 0, \quad x \in J.$$

*Proof of (i°).*  $\{\tau^*(J_\epsilon) \leq t\} = \{\tau_0^*(J_\epsilon) \leq T(t)\}$ . In the latter set, by taking inverses we get

$$\{\tau_0^*(J_\epsilon) \leq T(t)\} = \left\{ \int_0^{\tau_0^*(J_\epsilon)} V(X_0(\xi)) d\xi \leq t \right\}.$$

But letting  $M = \inf_I V(y)$ , we find that, denoting  $J_\epsilon^0 = (-\epsilon, \epsilon)$ ,

$$P_x \left( \int_0^{\tau_0^*(J_\epsilon)} V(X_0(\xi)) d\xi \leq t \right) \leq P_x(M\tau_0^*(J_\epsilon) \leq t) = P_0(\tau_0^*(J_\epsilon^0) \leq t/M).$$

Since  $P_x(|X(t) - x| \geq \epsilon) \leq P_x(\tau^*(J_\epsilon) \leq t)$ , condition (i) of the diffusion definition 16.76 is satisfied.

To do (ii) and (iii): By (i°) it is sufficient to prove

$$\frac{1}{t} \int_{\{t < \tau^*(J_\epsilon)\}} (X(t) - x) dP_x \xrightarrow{bp} 0, \quad \frac{1}{t} \int_{\{t < \tau^*(J_\epsilon)\}} (X(t) - x)^2 dP_x \xrightarrow{bp} \frac{1}{V(x)},$$

for  $x$  in finite  $J$ ,  $J \subset \text{int}(F)$ .

Let  $\tau_\epsilon^* = \min(t, \tau^*(J_\epsilon))$ . Then we will show that

$$(ii^\circ) \quad E_x X(\tau_\epsilon^*) = x, \quad E_x X^2(\tau_\epsilon^*) = x^2 + E_x T(\tau_\epsilon^*).$$

*Proof of (ii°).* Use (16.52); that is (dropping the  $\epsilon$ ),

$$\begin{aligned} \tau_0^*(J) &= T(\tau^*(J)), \\ X(\tau^*) &= X_0(T(\tau^*)) = X_0(\min(T(t), \tau_0^*(J))). \end{aligned}$$

But  $\tau_0^* = \min(T(t), \tau_0^*(J))$  is a stopping time for Brownian motion. Since  $X_0(t)$ ,  $X_0^2(t) - t$  are martingales, (ii°) follows.

Use (ii°) as follows:

$$\int_{\{t < \tau^*(J_\epsilon)\}} X(t) dP_x = \int_{\{t < \tau^*(J_\epsilon)\}} X(\tau_\epsilon^*) dP_x = x - \int_{\{t \geq \tau^*(J_\epsilon)\}} X(\tau_\epsilon^*) dP_x$$

Hence, we get the identity

$$\int_{\{t < \tau^*(J_\epsilon)\}} (X(t) - x) dP_x = - \int_{\{t \geq \tau^*(J_\epsilon)\}} (X(\tau^*(J_\epsilon)) - x) dP_x.$$

This latter integral is in absolute value less than  $\epsilon P_x(\tau^*(J_\epsilon) \leq t)$ . Apply (i°) now.

For (iii), write

$$\begin{aligned} \int_{\{t < \tau^*(J_\epsilon)\}} (X(t) - x)^2 dP_x &= \int_{\{t < \tau^*(J_\epsilon)\}} (X(\tau_\epsilon^*) - x)^2 dP_x \\ &= E_x(X(\tau_\epsilon^*) - x)^2 - \int_{\{t \geq \tau^*(J_\epsilon)\}} (X(\tau_\epsilon^*) - x)^2 dP_x. \end{aligned}$$

By (ii°),

$$E_x(X(\tau_\epsilon^*) - x)^2 = E_x T(\tau_\epsilon^*).$$

The remaining integral above is bounded by  $\epsilon^2 P_x(\tau^*(J_\epsilon) \leq t)$ , hence gives no trouble. To evaluate  $E_x T(\tau_\epsilon^*)$ , write

$$t \geq \int_0^{T(\tau_\epsilon^*)} V(X_0(\xi)) d\xi \geq M T(\tau_\epsilon^*),$$

so that

$$\frac{1}{t} E_x T(\tau_\epsilon^*) \leq \frac{1}{M},$$

for all  $t$  and  $x \in J$ . Further, since  $T(t) \rightarrow 0$  as  $t \rightarrow 0$ ,

$$\frac{t}{T(t)} = \frac{1}{T(t)} \int_0^{T(t)} V(X_0(\xi)) d\xi \rightarrow V(X_0(0)) \quad \text{a.s.}$$

as  $t \rightarrow 0$ . Since  $T(\tau_\epsilon^*)/t$  is bounded by  $M$ , the bounded convergence theorem can be applied to get

$$\lim_{t \rightarrow 0} \frac{1}{t} E_x T(\tau_\epsilon^*) = \frac{1}{V(x)}$$

for every  $x \in J$ .

These results help us to characterize diffusions. For example, now we know that of all Feller processes on natural scale, the diffusion processes are those such that  $dm \ll dx$  on  $\text{int}(F)$  and  $dm/dx$  has a continuous version, positive on  $\text{int}(F)$ . A nonscaled diffusion has the same speed measure property, plus a scale function with nonvanishing first derivative and continuous second derivative.

**Problem 19.** Show that for a diffusion, the functions  $\varphi_+$ ,  $\varphi_-$  of Section 10 are solutions in  $\text{int}(F)$  of

$$\frac{1}{2} \sigma^2 \frac{d^2 \varphi}{dx^2} + \mu \frac{d\varphi}{dx} = \lambda \varphi.$$

## NOTES

Most of the material in this chapter was gleaned from two recent books on this subject, Dynkin [44], of which an excellent English translation was published in 1965, and Ito, McKean [76, 1965].

The history of the subject matter is very recent. The starting point was a series of papers by Feller, beginning in 1952 [57], which used the semigroup approach. See also [58]. Dynkin introduced the idea of the characteristic operator in [43, 1955], and subsequently developed the theory using the associated concepts. The idea of random time substitutions was first exploited in this context by Volkonskiĭ [138, 1958]. The construction of the general process using local time was completed by Ito and McKean [76].

The material as it stands now leaves one a little unhappy from the pedagogic point of view. Some hoped-for developments would be: (1) a simple proof of the local time theorem (David Freedman has shown me a simple proof of all the results of the theorem excepting the continuity of  $l^*(t, y)$ ); (2) a direct proof of the unique determination of the process by scale function and speed measure to replace the present detour by means of the characteristic operator; (3) a simplification of the proofs of the existence of densities and smoothness properties for the transition probabilities.

Charles Stone has a method of getting Feller processes as limits of birth and death processes which seems to be a considerable simplification both conceptually and mathematically. Most of this is unpublished. However, see [131] for a bit of it.

The situation in two or more dimensions is a wilderness. The essential property in one dimension that does not generalize is that if a path-continuous process goes from  $x$  to  $y$ , then it has to pass through all the points between  $x$  and  $y$ . So far, the most powerful method for dealing with diffusions in any number of dimensions is the use of stochastic integral equations (see Doob [39, Chap. VI], Dynkin [44, Chap. XI]) initiated by Ito. The idea here is to attempt a direct integration to solve the equation

$$\Delta Y = \mu(Y) \Delta t + \sigma^2(Y) \Delta X$$

and its multidimensional analogs.

## APPENDIX

### ON MEASURE AND FUNCTION THEORY

The purpose of this appendix is to give a brief review, with very few proofs, of some of the basic theorems concerning measure and function theory. We refer for the proofs to Halmos [64] by page number.

#### 1. MEASURES AND THE EXTENSION THEOREM

For  $\Omega$  a set of points  $\omega$ , define

**Definition A.1.** A class  $\mathcal{F}$  of subsets of  $\Omega$  is a field if  $A, B \in \mathcal{F}$  implies  $A^c$ ,  $A \cup B$ ,  $A \cap B$  are in  $\mathcal{F}$ . The class  $\mathcal{F}$  is  $\sigma$ -field if it is a field, and if, in addition,  $A_n \in \mathcal{F}$ ,  $n = 1, 2, \dots$  implies  $\bigcup_1^\infty A_n \in \mathcal{F}$ .

**Notation A.2.** We will use  $A^c$  for the complement of  $A$ ,  $A - B$  for  $A \cap B^c$ ,  $\emptyset$  for the empty set,  $A \Delta B$  for the symmetric set difference

$$(A - B) \cup (B - A).$$

Note

**Proposition A.3.** For any class  $\mathcal{C}$  of subsets of  $\Omega$ , there is a smallest field of subsets, denoted  $\mathcal{F}_0(\mathcal{C})$ , and a smallest  $\sigma$ -field of subsets, denoted  $\mathcal{F}(\mathcal{C})$ , containing all the sets in  $\mathcal{C}$ .

*Proof.* The class of all subsets of  $\Omega$  is a  $\sigma$ -field containing  $\mathcal{C}$ . Let  $\mathcal{F}(\mathcal{C})$  be the class of sets  $A$  such that  $A$  is in every  $\sigma$ -field that contains  $\mathcal{C}$ . Check that  $\mathcal{F}(\mathcal{C})$  so defined is a  $\sigma$ -field and that if  $\mathcal{F}$  is a  $\sigma$ -field,  $\mathcal{C} \subset \mathcal{F}$ , then  $\mathcal{F}(\mathcal{C}) \subset \mathcal{F}$ . For fields a finite construction will give  $\mathcal{F}_0(\mathcal{C})$ .

If  $A_n$  is a sequence of sets such that  $A_n \subset A_{n+1}$ ,  $n = 1, 2, \dots$  and  $A = \bigcup A_n$ , write  $A_n \uparrow A$ . Similarly, if  $A_{n+1} \subset A_n$ ,  $A = \bigcap A_n$ , write  $A_n \downarrow A$ . Define a monotone class of subsets  $\mathcal{C}$  by: If  $A_n \in \mathcal{C}$ , and  $A_n \uparrow A$  or  $A_n \downarrow A$  then  $A \in \mathcal{C}$ .

**Monotone Class Theorem A.4** (Halmos, p. 27). The smallest monotone class of sets containing a field  $\mathcal{F}_0$  is  $\mathcal{F}(\mathcal{F}_0)$ .

**Definition A.5.** A finitely additive measure  $\mu$  on a field  $\mathcal{F}$  is a real-valued (including  $+\infty$ ), nonnegative function with domain  $\mathcal{F}$  such that for  $A, B \in \mathcal{F}$ ,  $A \cap B = \emptyset$ ,

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

This extends to: If  $A_1, \dots, A_n \in \mathcal{F}$  are pairwise disjoint,  $A_i \cap A_j = \emptyset$ ,  $i \neq j$ , then

$$\mu\left(\bigcup_1^n A_j\right) = \sum_1^n \mu(A_j).$$

Whether or not the sets  $A_1, \dots, A_n \in \mathcal{F}$  are disjoint,

$$\mu\left(\bigcup_1^n A_j\right) \leq \sum_1^n \mu(A_j).$$

**Definition A.6.** A  $\sigma$ -additive measure (or just measure) on a  $\sigma$ -field  $\mathcal{F}$  is a real-valued ( $+\infty$  included), nonnegative function with domain  $\mathcal{F}$  such that for  $A_1, \dots \in \mathcal{F}$ ,  $A_i \cap A_j = \emptyset$ ,  $i \neq j$ ,

$$\mu\left(\bigcup_j A_j\right) = \sum_j \mu(A_j).$$

We want some finiteness:

**Definition A.7.** A measure (finitely or  $\sigma$ -additive) on a field  $\mathcal{F}_0$  is  $\sigma$ -finite if there are sets  $A_k \in \mathcal{F}_0$  such that  $\bigcup_k A_k = \Omega$  and for every  $k$ ,  $\mu(A_k) < \infty$ .

We restrict ourselves henceforth to  $\sigma$ -finiteness! The extension problem for measures is: Given a finitely additive measure  $\mu_0$  on a field  $\mathcal{F}_0$ , when does there exist a measure  $\mu$  on  $\mathcal{F}(\mathcal{F}_0)$  agreeing with  $\mu_0$  on  $\mathcal{F}_0$ ? A measure has certain continuity properties:

**Proposition A.8.** Let  $\mu$  be a measure on the  $\sigma$ -field  $\mathcal{F}$ . If  $A_n \downarrow A$ ,  $A_n \in \mathcal{F}$ , and if  $\mu(A_n) < \infty$  for some  $n$ , then

$$\lim_n \mu(A_n) = \mu(A).$$

Also, if  $A_n \uparrow A$ ,  $A_n \in \mathcal{F}$ , then

$$\lim_n \mu(A_n) = \mu(A).$$

This is called continuity from above and below. Certainly, if  $\mu_0$  is to be extended, then the minimum requirement needed is that  $\mu_0$  be continuous on its domain. Call  $\mu_0$  continuous from above at  $\emptyset$  if whenever  $A_n \in \mathcal{F}_0$ ,  $A_n \downarrow \emptyset$ , and  $\mu_0(A_n) < \infty$  for some  $n$ , then

$$\lim_n \mu_0(A_n) = 0.$$

**Carathéodory Extension Theorem A.9.** If  $\mu_0$  on  $\mathcal{F}_0$  is continuous from above at  $\emptyset$ , then there is a unique measure  $\mu$  on  $\mathcal{F}(\mathcal{F}_0)$  agreeing with  $\mu_0$  on  $\mathcal{F}_0$  (see Halmos, p. 54).

**Definition A.10.** A measure space is a triple  $(\Omega, \mathcal{F}, \mu)$  where  $\mathcal{F}, \mu$  are a  $\sigma$ -field and measure. The completion of a measure space, denoted by  $(\Omega, \bar{\mathcal{F}}, \bar{\mu})$ , is gotten by defining  $A \in \bar{\mathcal{F}}$  if there are sets  $A_1, A_2$  in  $\mathcal{F}$ ,  $A_1 \subset A \subset A_2$  and  $\mu(A_2 - A_1) \triangleq 0$ . Then define  $\bar{\mu}(A) = \mu(A_1)$ .

$\bar{\mathcal{F}}$  is the largest  $\sigma$ -field for which unique extension under the hypothesis of A.9 holds. That is,  $\bar{\mu}$  is the only measure on  $\bar{\mathcal{F}}$  agreeing with  $\mu_0$  on  $\mathcal{F}_0$  and

**Proposition A.11.** Let  $B \notin \bar{\mathcal{F}}, \mathcal{F}_1$  the smallest  $\sigma$ -field containing both  $B$  and  $\bar{\mathcal{F}}$ . Then there is an infinity of measures on  $\mathcal{F}_1$  agreeing with  $\mu_0$  on  $\mathcal{F}_0$ . (See Halmos, p. 55 and p. 71, Problem 3).

Note that  $\mathcal{F}(\mathcal{F}_0)$  depends only on  $\mathcal{F}_0$  and not on  $\mu_0$ , but that  $\bar{\mathcal{F}}(\mathcal{F}_0)$  depends on  $\mu_0$ . The measure  $\mu$  on  $\mathcal{F}(\mathcal{F}_0)$ , being a unique extension, must be approximable in some sense by  $\mu_0$  on  $\mathcal{F}_0$ . One consequence of the extension construction is

**Proposition A.12** (Halmos, p. 56). For every  $A \in \mathcal{F}(\mathcal{F}_0)$ , and  $\epsilon > 0$ , there is a set  $A_0 \in \mathcal{F}_0$  such that

$$\mu(A \triangle A_0) \leq \epsilon.$$

We will designate a space  $\Omega$  and a  $\sigma$ -field  $\mathcal{F}$  of subsets of  $\Omega$  as a *measurable space*  $(\Omega, \mathcal{F})$ . If  $F \subset \Omega$ , denote by  $\mathcal{F}(F)$  the  $\sigma$ -field of subsets of  $F$  of the form  $A \cap F$ ,  $A \in \mathcal{F}$ , and take the complement relative to  $F$ .

Some important measurable spaces are

$R^{(1)}$	the real line
$\mathcal{B}_1$	the smallest $\sigma$ -field containing all intervals
$R^{(k)}$	$k$ -dimensional Euclidean space
$\mathcal{B}_k$	the smallest $\sigma$ -field containing all $k$ -dimensional rectangles
$R^{(\infty)}$	the space of all infinite sequences $(x_1, x_2, \dots)$ of real numbers
$\mathcal{B}_\infty$	the smallest $\sigma$ -field containing all sets of the form $\{(x_1, x_2, \dots); x_1 \in I_1, \dots, x_n \in I_n\}$ for any $n$ where $I_1, \dots, I_n$ are any intervals
$R^I$	the space of all real-valued functions $x(t)$ on the interval $I \subset R^{(1)}$
$\mathcal{B}_I$	the smallest $\sigma$ -field containing all sets of the form $\{x(\cdot) \in R^I; x(t_1) \in I_1, \dots, x(t_n) \in I_n\}$ for any $t_1, \dots, t_n \in I$ and intervals $I_1, \dots, I_n$

**Definition A.13.** The class of all finite unions of disjoint intervals in  $R^{(1)}$  is a field. Take  $\mu_0$  on this field to be length. Then  $\mu_0$  is continuous from above at  $\emptyset$  (Halmos, pp. 34 ff.). The extension of length to  $\mathcal{B}_1$  is Lebesgue measure, denoted by  $l$  or by  $dx$ .

Henceforth, if we have a measure space  $(\Omega, \mathcal{F}, \mu)$  and a statement holds for all  $\omega \in \Omega$  with the possible exception of  $\omega \in A$ , where  $\mu(A) = 0$ , we say that the statement holds *almost everywhere* (a.e.)

## 2. MEASURABLE MAPPINGS AND FUNCTIONS

**Definition A.14.** Given two spaces  $\Omega, R$ , and a mapping  $X(\omega): \Omega \rightarrow R$ , the inverse image of a set  $B \subset R$  is defined as

$$X^{-1}B = \{\omega \in \Omega; X(\omega) \in B\}.$$

Denote this by  $\{X \in B\}$ . The taking of inverse images preserves all set operations; that is,

$$\begin{aligned}\left\{X \in \bigcup_{\lambda} B_{\lambda}\right\} &= \bigcup_{\lambda} \{X \in B_{\lambda}\}, \\ \left\{X \in \bigcap_{\lambda} B_{\lambda}\right\} &= \bigcap_{\lambda} \{X \in B_{\lambda}\}, \\ \{X \in B^c\} &= \{X \in B\}^c.\end{aligned}$$

**Definition A.15.** Given two measurable spaces  $(\Omega, \mathcal{F})$ ,  $(R, \mathcal{B})$ . A mapping  $X: \Omega \rightarrow R$  is called measurable if the inverse of every set in  $\mathcal{B}$  is in  $\mathcal{F}$ .

**Proposition A.16** (See 2.29). Let  $\mathcal{C} \subset \mathcal{B}$  such that  $\mathcal{F}(\mathcal{C}) = \mathcal{B}$ . Then  $X: \Omega \rightarrow R$  is measurable if the inverse of every set in  $\mathcal{C}$  is in  $\mathcal{F}$ .

**Definition A.17.**  $X: \Omega \rightarrow R^{(1)}$  will be called a measurable function if it is a measurable map from  $(\Omega, \mathcal{F})$  to  $(R^{(1)}, \mathcal{B}_1)$ .

From A.16 it is sufficient that  $\{X < x\} \in \mathcal{F}$  for all  $x$  in a set dense in  $R^{(1)}$ . Whether or not a function is measurable depends on both  $\Omega$  and  $\mathcal{F}$ . Refer therefore to measurable functions on  $(\Omega, \mathcal{F})$  as  $\mathcal{F}$ -measurable functions.

**Proposition A.18.** The class of  $\mathcal{F}$ -measurable functions is closed under pointwise convergence. That is, if  $X_n(\omega)$  are each  $\mathcal{F}$ -measurable, and  $\lim_n X_n(\omega)$  exists for every  $\omega$ , then  $X(\omega) = \lim_n X_n(\omega)$  is  $\mathcal{F}$ -measurable.

*Proof.* Suppose  $X_n(\omega) \downarrow X(\omega)$  for each  $\omega$ ; then  $\{X < x\} = \bigcup_n \{X_n < x\}$ . This latter set is in  $\mathcal{F}$ . In general, if  $X_n(\omega) \rightarrow X(\omega)$ , take

$$Y_n(\omega) = \sup_{m \geq n} X_m(\omega).$$

Then  $\{Y_n > y\} = \bigcup_{m \geq n} \{X_m > y\}$  which is in  $\mathcal{F}$ . Then  $Y_n$  is  $\mathcal{F}$ -measurable and  $Y_n \downarrow X$ .

Define an *extended measurable function* as a function  $X(\omega)$  which takes values in the extended real line  $R^{(1)} \cup \{\infty\}$  such that  $\{X \in B\} \in \mathcal{F}$  for every  $B \in \mathcal{B}_1$ .

By the argument above, if  $X_n$  are  $\mathcal{F}$ -measurable, then  $\overline{\lim} X_n$ ,  $\underline{\lim} X_n$  are extended  $\mathcal{F}$ -measurable, hence the set

$$\{\lim X_n \text{ exists}\} = \{\underline{\lim} X_n = \overline{\lim} X_n\} \cap \{|\underline{\lim} X_n| < \infty\}$$

is in  $\mathcal{F}$ .

**Proposition A.19** (See 2.31). *If  $X$  is a measurable mapping from  $(\Omega, \mathcal{F})$  to  $(R, \mathcal{B})$  and  $\varphi$  is a  $\mathcal{B}$ -measurable function, then  $\varphi(X)$  is an  $\mathcal{F}$ -measurable function.*

The *set indicator* of a subset  $A \subset \Omega$  is the function

$$\chi_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in A^c. \end{cases}$$

A *simple function* is any finite linear combination of set indicators,

$$g(\omega) = \sum_{k=1}^n \alpha_k \chi_{A_k}(\omega)$$

of sets  $A_k \in \mathcal{F}$ .

**Proposition A.20.** *The class of  $\mathcal{F}$ -measurable functions is the smallest class of functions containing all simple functions and closed under pointwise convergence.*

*Proof.* For any  $n \geq 0$  and  $X(\omega)$  a measurable function, define sets  $A_k = \{X \in [k/n, (k+1)/n)\}$  and consider

$$X_n = \sum_{k=-n^2}^{n^2} \frac{k}{n} \chi_{A_k}(\omega).$$

Obviously  $X_n \rightarrow X$ .

For any measurable mapping  $X$  from  $(\Omega, \mathcal{F})$  to  $(R, \mathcal{B})$ , denote by  $\mathcal{F}(X)$  the  $\sigma$ -field of inverse images of sets in  $\mathcal{B}$ . Now we prove 4.9.

**Proposition A.21.** *If  $Z$  is an  $\mathcal{F}(X)$ -measurable function, then there is a  $\mathcal{B}$ -measurable function  $\theta$  such that  $Z = \theta(X)$ .*

*Proof.* Consider the class of functions  $\varphi(X)$ ,  $X$  fixed, as  $\varphi$  ranges over the  $\mathcal{B}$ -measurable functions. Any set indicator  $\chi_A(\omega)$ ,  $A \in \mathcal{F}(X)$ , is in this class, because  $A = \{X \in B\}$  for some  $B \in \mathcal{B}$ . Hence

$$\chi_A(\omega) = \chi_B(X).$$

Now the class is closed under addition, so by A.20 it is sufficient to show it closed under pointwise convergence. Let  $\varphi_n(X) \rightarrow Y$ ,  $\varphi_n$   $\mathcal{B}$ -measurable.



Let  $B = \{\lim_n \varphi_n \text{ exists}\}$ . Then  $B \in \mathfrak{B}$ , and  $\Omega = \{X \in B\}$ . Define

$$\varphi = \begin{cases} \lim_n \varphi_n, & \text{on } B \\ 0, & \text{on } B^c. \end{cases}$$

Obviously, then  $Y = \varphi(X)$ .

We modify the proof of A.20 slightly to get 2.38;

**Proposition A.22.** Consider a class  $\mathcal{L}$  of  $\mathcal{F}$ -measurable functions having the properties

- i)  $X, Y \in \mathcal{L}, \alpha, \beta \geq 0 \Rightarrow \alpha X + \beta Y \in \mathcal{L}$ ,
- ii)  $X_n \in \mathcal{L}, X_n \uparrow X \Rightarrow X \in \mathcal{L}$ ,
- iii) for every  $A \in \mathcal{F}, \chi_A(\omega) \in \mathcal{L}$ ;

then  $\mathcal{L}$  includes all nonnegative  $\mathcal{F}$ -measurable functions.

*Proof.* For  $X \geq 0$ ,  $\mathcal{F}$ -measurable, let  $X_n = \sum_{k=0}^{n^2} k/n \chi_{A_k}(\omega)$  where  $A_k = \{X \in [k/n, (k+1)/n)\}$ . Then certainly  $X_n \in \mathcal{L}$ , and  $X_n \uparrow X$  if we take  $n'$  the subsequence  $\{2^m\}$ .

### 3. THE INTEGRAL

Take  $(\Omega, \mathcal{F}, \mu)$  to be a measure space. Let  $X(\omega) \geq 0$  be a nonnegative  $\mathcal{F}$ -measurable function. To define the integral of  $X$  let  $X_n \geq 0$  be simple functions such that  $X_n \uparrow X$ .

**Definition A.23.** The integral  $\int X_n d\mu$  of the nonnegative simple function  $X_n = \sum \alpha_k \chi_{A_k}(\omega)$ ,  $\alpha_k \geq 0$ , is defined by  $\sum \alpha_k \mu(A_k)$ .

For  $X_n \uparrow X$ , it is easy to show that  $\int X_{n+1} d\mu \geq \int X_n d\mu \geq 0$ .

**Definition A.24.** Define  $\int X d\mu$  as  $\lim_n \int X_n d\mu$ . Furthermore, the value of this limit is the same for all sequences of nonnegative simple functions converging up to  $X$  (Halmos, p. 101).

Note that this limit may be infinite. For any  $\mathcal{F}$ -measurable function  $X$ , suppose that  $\int |X| d\mu < \infty$ . In this case define  $\mathcal{F}$ -measurable functions

$$X^+(\omega) = \begin{cases} X(\omega), & X(\omega) \geq 0, \\ 0, & X(\omega) < 0, \end{cases}$$

$$X^-(\omega) = \begin{cases} -X(\omega), & X(\omega) \leq 0, \\ 0, & X(\omega) > 0. \end{cases}$$

Note  $|X| = X^+ + X^-$ .

**Definition A.25.** If  $\int |X| d\mu < \infty$ , define

$$\int X d\mu = \int X^+ d\mu - \int X^- d\mu;$$

we may sometimes use the notation

$$\int X(\omega) \mu(d\omega).$$

The elementary properties of the integral are: If the integrals of  $X$  and  $Y$  exist,

$$\text{i) } X \geq Y \Rightarrow \int X d\mu \geq \int Y d\mu,$$

$$\text{ii) } \int (\alpha X + \beta Y) d\mu = \alpha \int X d\mu + \beta \int Y d\mu,$$

$$\text{iii) } A, B \in \mathcal{F}, A \cap B = \emptyset \Rightarrow \int_{A \cup B} X d\mu = \int_A X d\mu + \int_B X d\mu.$$

Some nonelementary properties begin with

**Monotone Convergence Theorem A.26** (Halmos, p. 112). For  $X_n \geq 0$  non-negative  $\mathcal{F}$ -measurable functions,  $X_n \uparrow X$ , then

$$\lim \int X_n d\mu = \int X d\mu.$$

From this comes the

**Fatou Lemma A.27.** If  $X_n \geq 0$ , then

$$\int \underline{\lim} X_n d\mu \leq \underline{\lim} \int X_n d\mu.$$

*Proof.* To connect up with A.26 note that for  $X_1, \dots, X_n$  arbitrary, non-negative  $\mathcal{F}$ -measurable functions,

$$\int \inf(X_1, \dots, X_n) d\mu \leq \int X_k d\mu, \quad k = 1, \dots, n.$$

Hence, by taking limits

$$\int \inf_n X_n d\mu \leq \inf_n \int X_n d\mu.$$

Let  $Y_n = \inf_{m \geq n} X_m$ ; then

$$\int Y_n d\mu \leq \inf_{m \geq n} \int X_m d\mu.$$

Since  $Y_n \geq 0$ , and  $Y_n \uparrow \underline{\lim} X_n$ , apply A.26 to complete the proof.

Another useful convergence result is:

**Bounded Convergence Theorem A.28** (2.44). Let  $X_n \rightarrow X$  pointwise, where the  $X_n$  are  $\mathcal{F}$ -measurable functions such that there is an  $\mathcal{F}$ -measurable function  $Z$  with  $|X_n| \leq Z$ , all  $n$ ,  $\omega$ , and  $\int Z d\mu < \infty$ . Then (see Halmos, p. 110)

$$\lim_n \int X_n d\mu = \int X d\mu.$$

From these convergence theorems can be deduced the  $\sigma$ -additivity of an integral: For  $\{B_n\}$  disjoint,  $B_n \in \mathcal{F}$ , and  $\int |X| d\mu < \infty$ ,

$$\int_{\cup B_n} X d\mu = \sum \int_{B_n} X d\mu.$$

Also, if  $\int |X| d\mu < \infty$ , then the integral is absolutely continuous. That is, for every  $\epsilon > 0$ , there exists a  $\delta > 0$  such that if  $A \in \mathcal{F}$  and  $\mu(A) \leq \delta$ , then

$$\int_A X d\mu \leq \epsilon.$$

#### 4. ABSOLUTE CONTINUITY AND THE RADON-NIKODYM THEOREM

Consider a measurable space  $(\Omega, \mathcal{F})$  and two measures  $\mu, \nu$  on  $\mathcal{F}$ .

**Definition A.29.** Say that  $\nu$  is absolutely continuous with respect to  $\mu$ , denoted  $\nu \ll \mu$ , if  $A \in \mathcal{F}$ ,  $\mu(A) = 0 \Rightarrow \nu(A) = 0$ .

Call two measurable functions  $X_1, X_2$  equivalent if  $\mu(\{X_1 \neq X_2\}) = 0$ . Then

**Radon-Nikodym Theorem A.30** (Halmos, p. 128). If  $\nu \ll \mu$ , then there exists a nonnegative  $\mathcal{F}$ -measurable function  $X$  determined up to equivalence, such that for any  $A \in \mathcal{F}$ ,

$$\nu(A) = \int_A X d\mu.$$

Another way of denoting this is to say that the Radon derivative of  $\nu$  with respect to  $\mu$  exists and equals  $X$ ; that is,

$$\frac{d\nu}{d\mu} = X.$$

The opposite of continuity is

**Definition A.31.** Say that  $\nu$  is singular with respect to  $\mu$ , written  $\mu \perp \nu$  if there exists  $A \in \mathcal{F}$  such that

$$\mu(A) = 0, \quad \nu(A^c) = 0.$$

**Lebesgue Decomposition Theorem A.32** (Halmos, p. 134). For any two measures  $\mu, \nu$  on  $\mathcal{F}$ ,  $\nu$  can be decomposed into two measures,  $\nu_c, \nu_s$ , in the sense that for every  $A \in \mathcal{F}$ ,

$$\nu(A) = \nu_c(A) + \nu_s(A)$$

and  $\nu_c \ll \mu, \nu_s \perp \mu$ .

For a  $\sigma$ -finite measure the set of points  $\{\omega; \mu(\{\omega\}) > 0\}$  is at most countable. Call  $\nu$  a *point measure* if there is a countable set  $G = \{\omega_j\}$  such that for every  $A \in \mathcal{F}$ ,

$$\nu(A) = \nu(A \cap G).$$

Obviously, any measure  $\nu$  may be decomposed into  $\nu_1 + \nu_p$ , where  $\nu_p$  is a point measure and  $\nu_1$  assigns mass zero to any one-point set. Hence, on  $(R^{(1)}, \mathcal{B}_1)$  we have the special case of A.32.

**Corollary A.33.** A measure  $\nu$  on  $\mathcal{B}_1$  can be written as

$$\nu = \nu_c + \nu_s + \nu_p,$$

where  $\nu_p$  is a point measure,  $\nu_s \perp l$  but  $\nu_s$  assigns zero mass to any one-point sets, and  $\nu_c \ll l$ . [Recall  $l$  is Lebesgue measure.]

## 5. CROSS-PRODUCT SPACES AND THE FUBINI THEOREM

**Definition A.34.** Given two spaces  $\Omega_1, \Omega_2$ , their cross product  $\Omega_1 \times \Omega_2$  is the set of all ordered pairs  $\{(\omega_1, \omega_2); \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$ . For measurable spaces  $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ ,  $\mathcal{F}_1 \times \mathcal{F}_2$  is the smallest  $\sigma$ -field containing all sets of the form

$$\{(\omega_1, \omega_2); \omega_1 \in A_1, \omega_2 \in A_2\},$$

where  $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$ . Denote this set by  $A_1 \times A_2$ .

For a function  $X(\omega_1, \omega_2)$  on  $\Omega_1 \times \Omega_2$ , its *section* at  $\omega_1$  is the function on  $\Omega_2$  gotten by holding  $\omega_1$  constant and letting  $\omega_2$  be the variable. Similarly, if  $A \subset \Omega_1 \times \Omega_2$ , its *section* at  $\omega_1$  is defined as  $\{\omega_2; (\omega_1, \omega_2) \in A\}$ .

**Theorem A.35** (Halmos, pp. 141ff.). Let  $X$  be an  $\mathcal{F}_1 \times \mathcal{F}_2$ -measurable function; then every section of  $X$  is an  $\mathcal{F}_2$ -measurable function. If  $A \in \mathcal{F}_1 \times \mathcal{F}_2$ , every section of  $A$  is in  $\mathcal{F}_2$ .

If we have measures  $\mu_1$  on  $\mathcal{F}_1, \mu_2$  on  $\mathcal{F}_2$ , then

**Theorem A.36** (Halmos, p. 144). There is a unique measure  $\mu_1 \times \mu_2$  on  $\mathcal{F}_1 \times \mathcal{F}_2$  such that for every  $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$ ,

$$\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2).$$

This is called the cross-product measure.

**Fubini Theorem A.37** (Halmos, p. 148). Let  $X$  be  $\mathcal{F}_1 \times \mathcal{F}_2$ -measurable, and

$$\int |X| d(\mu_1 \times \mu_2) < \infty.$$

Then

$$\int X(\omega_1, \omega_2) \mu_1(d\omega_1), \quad \int X(\omega_1, \omega_2) \mu_2(d\omega_2)$$

are respectively  $\mathcal{F}_2$ - and  $\mathcal{F}_1$ -measurable functions, which may be infinite on sets of measure zero, but whose integrals exist. And

$$\begin{aligned} \int X d(\mu_1 \times \mu_2) &= \int \left( \int X d\mu_1 \right) d\mu_2 \\ &= \int \left( \int X d\mu_2 \right) d\mu_1. \end{aligned}$$

**Corollary A.38.** If  $A \in \mathcal{F}_1 \times \mathcal{F}_2$  and  $\mu_1 \times \mu_2(A) = 0$ , then almost every section of  $A$  has  $\mu_2$  measure zero.

This all has fairly obvious extensions to finite cross products  $\Omega_1 \times \cdots \times \Omega_n$ .

## 6. THE $L_r(\mu)$ SPACES

These are some well-known results. Let  $(\Omega, \mathcal{F}, \mu)$  be a measure space and the functions  $X, Y$  be  $\mathcal{F}$ -measurable:

**Schwarz Inequality A.39.** If  $\int |X|^2 d\mu$  and  $\int |Y|^2 d\mu$  are finite, then so is  $\int |XY| d\mu$  and

$$\left( \int XY d\mu \right)^2 \leq \left( \int |X|^2 d\mu \right) \left( \int |Y|^2 d\mu \right).$$

**Definition A.40.**  $L_r(\mu)$ ,  $r > 0$ , is the class of all  $\mathcal{F}$ -measurable functions  $X$  such that  $\int |X|^r d\mu < \infty$ .

**Completeness Theorem A.41** (Halmos, p. 107). If  $X_n \in L_r(\mu)$  and

$$\int |X_n - X_m|^r d\mu \rightarrow 0$$

as  $m, n \rightarrow \infty$  in any way, then there is a function  $X \in L_r(\mu)$  such that

$$\int |X_n - X|^r d\mu \rightarrow 0.$$

## 7. TOPOLOGICAL MEASURE SPACES

In this section, unless otherwise stated, assume that  $\Omega$  has a metric under which it is a separable metric space. Let  $\mathcal{C}$  be the class of all open sets in  $\Omega$ . A measure  $\mu$  on  $\mathcal{F}$  is called *inner regular* if for any  $A \in \mathcal{F}$

$$\mu(A) = \sup_C \mu(C),$$

where the sup is over all compact sets  $C \subset A$ ,  $C \in \mathcal{F}$ . It is called *outer regular* if

$$\mu(A) = \inf_0 \mu(O),$$

where the inf is over all open sets  $O$  such that  $A \subset O$ , and  $O \in \mathcal{F}$ .

**Theorem A.42** (Follows from Halmos, p. 228). *Any measure on  $\mathcal{F}(\mathbb{C})$  is both inner and outer regular.*

**Theorem A.43** (Follows from Halmos, p. 240). *The class of  $\mathcal{F}(\mathbb{C})$ -measurable functions is the smallest class of functions containing all continuous functions on  $\Omega$  and closed under pointwise convergence.*

**Theorem A.44** (Halmos, p. 241). *If  $\int |X| d\mu < \infty$  for  $X$   $\mathcal{F}(\mathbb{C})$ -measurable, then for any  $\epsilon > 0$  there is a continuous function  $\varphi$  on  $\Omega$  such that*

$$\int |X - \varphi| d\mu \leq \epsilon.$$

**Definition A.45.** *Given any measurable space  $(\Omega, \mathcal{F})$  ( $\Omega$  not necessarily metric), it is called a *Borel space* if there is 1-1 mapping  $\varphi: \Omega \leftrightarrow E$  where  $E \in \mathcal{B}_1$  such that  $\varphi$  is  $\mathcal{F}$ -measurable, and  $\varphi^{-1}$  is  $\mathcal{B}_1(E)$ -measurable.*

**Theorem A.46.** *If  $\Omega$  is complete, then  $(\Omega, \mathcal{F}(\mathbb{C}))$  is a Borel space.*

We prove this in the case that  $(\Omega, \mathcal{F}(\mathbb{C}))$  is  $(R^{(\infty)}, \mathcal{B}_\infty)$ . Actually, since there is a 1-1 continuous mapping between  $R^{(1)}$  and  $(0, 1)$  it is sufficient to show that

**Theorem A.47.**  *$((0, 1)^{(\infty)}, \mathcal{B}_\infty(0, 1))$  is a Borel space.*

*Note.* Here  $(0, 1)^{(\infty)}$  is the set of all infinite sequences with coordinates in  $(0, 1)$  and  $\mathcal{B}_\infty(0, 1)$  means  $\mathcal{B}_\infty((0, 1))$ .

*Proof.* First we construct the mapping  $\Phi$  from  $(0, 1)$  to  $(0, 1)^{(\infty)}$ . Every number in  $(0, 1)$  has a unique binary expansion  $x = .x_1x_2 \cdots$  containing an infinite number of zeros. Consider the triangular array

$$\begin{array}{ccccccc} 1 & & & & & & \\ 2 & 3 & & & & & \\ 4 & 5 & 6 & & & & \\ 7 & 8 & 9 & 10 & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{array}$$

Let

$$\Phi(x) = (\Phi_1(x), \Phi_2(x), \dots),$$

where the  $n$ th coordinate is formed by going down the  $n$ th column of the array; that is,

$$\Phi_1(x) = .x_1x_2x_4x_7 \cdots,$$

$$\Phi_2(x) = .x_3x_6x_8 \cdots,$$

and so on. Conversely, if  $\mathbf{x} \in (0, 1)^{(\infty)}$ ,  $\mathbf{x} = (x^{(1)}, x^{(2)}, \dots)$ , expand every coordinate as the unique binary decimal having an infinite number of zeros, say

$$x^{(k)} = .x_1^{(k)}x_2^{(k)} \dots$$

and define  $\varphi(\mathbf{x})$  to be the binary decimal whose  $n$ th entry is  $x_j^{(k)}$  if  $n$  appears in the  $k$ th column  $j$  numbers down. That is,

$$\varphi(\mathbf{x}) = .x_1^{(1)}x_2^{(1)}x_1^{(2)}x_3^{(1)}x_2^{(2)}x_1^{(3)} \dots$$

Clearly,  $\Phi$  and  $\varphi$  are inverses of each other, so the mapping is 1-1 and onto. By 2.13, to show  $\Phi \mathcal{B}_1(0, 1)$ -measurable, it is sufficient to show that each  $\Phi_k(x)$  is  $\mathcal{B}_1(0, 1)$ -measurable. Notice that the coordinates  $x_1(x), x_2(x), \dots$  in the decimal expansion of  $x$  are measurable functions of  $x$ , continuous except at the points which have only a finite number of ones in their expansion (binary rationals). Furthermore, each  $\Phi_k(x)$  is a sum of these, for example,

$$\Phi_1(x) = \frac{x_1(x)}{2} + \frac{x_2(x)}{2^2} + \frac{x_4(x)}{2^3} + \dots$$

Therefore, every  $\Phi_k(x)$  is measurable. The proof that  $\varphi(\mathbf{x})$  is measurable similarly proceeds from the observation that  $x_j^{(k)}(\mathbf{x})$  is a measurable function of  $\mathbf{x}$ . Q.E.D.

To prove A.46 generally, see Sierpinski, [123a, p. 137], where it is proved that every complete separable metric space is homeomorphic to a subset of  $R^{(\infty)}$ . (See also p. 206.)

## 8. EXTENSION ON SEQUENCE SPACE

**A.48** (*Proof of Theorem 2.18*). Consider the class  $\mathcal{F}_0$  of all finite disjoint unions of finite dimensional rectangles. It is easily verified that  $\mathcal{F}_0$  is a field, and that

$$\mathcal{B}_\infty = \mathcal{F}(\mathcal{F}_0).$$

For any set  $A \in \mathcal{F}_0$ ,  $A = \bigcup_j S_j$ , where the  $S_j$  are disjoint rectangles, define

$$\hat{P}(A) = \sum_j \hat{P}(S_j).$$

There is a uniqueness problem: Suppose  $A$  can also be represented as  $\bigcup_k S'_k$ , where the  $S'_k$  are also disjoint. We need to know that

$$\sum_k \hat{P}(S'_k) = \sum_j \hat{P}(S_j).$$

Write

$$S'_k = S'_k \cap A = \bigcup_j (S'_k \cap S_j).$$

By part (b) of the hypothesis,

$$\hat{P}(S'_k) = \sum_j \hat{P}(S'_k \cap S_j);$$

so

$$\sum_k \hat{P}(S'_k) = \sum_{j,k} \hat{P}(S_j \cap S'_k).$$

By a symmetric argument

$$\sum_j \hat{P}(S_j) = \sum_{j,k} \hat{P}(S_j \cap S'_k).$$

Now we are in a position to apply the Carathéodory extension theorem, if we can prove that  $A_n \in \mathcal{F}_0$ ,  $A_n \downarrow \emptyset$  implies  $\hat{P}(A_n) \rightarrow 0$ . To do this, assume that  $\lim_n \hat{P}(A_n) = \delta > 0$ . By repeating some  $A_n$  in the sequence, if necessary, we can assume that

$$A_n = \{\mathbf{x}; (x_1, \dots, x_n) \in A_n^*\},$$

where  $A_n^*$  is a union of disjoint rectangles in  $R^{(n)}$ . By part (c) of the hypothesis we can find a set  $B_n^* \subset A_n^*$  so that  $B_n^*$  is a finite union of compact rectangles in  $R^{(n)}$ , and if

$$B_n = \{\mathbf{x}; (x_1, \dots, x_n) \in B_n^*\},$$

then

$$\hat{P}(A_n - B_n) \leq \frac{\delta}{2^{n+1}}.$$

Form the sets  $C_n = \bigcap_1^n B_k$ , and put

$$C_n = \{\mathbf{x}; (x_1, \dots, x_n) \in C_n^*\}.$$

Then, since the  $A_n$  are nonincreasing,

$$\begin{aligned} \hat{P}(A_n - C_n) &\leq \sum_{k=1}^n \hat{P}(A_n - B_k) \\ &\leq \sum_{k=1}^n \hat{P}(A_k - B_k) \\ &\leq \delta/2. \end{aligned}$$

The conclusion is that  $\lim_n \hat{P}(C_k) \geq \delta/2$ , and, of course,  $C_n \downarrow \emptyset$ . Take points  $\mathbf{x}^{(1)} \in C_1$ ,  $\mathbf{x}^{(2)} \in C_2$ , . . . ,

$$\mathbf{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots).$$

For every  $n$ ,

$$(x_1^{(n)}, \dots, x_n^{(n)}) \in C_n^*.$$

Take  $N_1$  any ordered infinite subsequence of integers such that  $x_1^{(n)} \rightarrow x_1 \in C_1^*$  as  $n$  runs through  $N_1$ . This is certainly possible since  $x_1^{(n)} \in C_1^*$  for all  $n$ .



Now take  $N_2 \subset N_1$  such that

$$(x_1^{(n)}, x_2^{(n)}) \rightarrow (x_1, x_2) \in C_2^*$$

as  $n$  runs through  $N_2$ . Continuing, we construct subsequences

$$N_1 \supset N_2 \supset N_3 \supset \dots$$

Let  $n_k$  be the  $k$ th member of  $N_k$ , then for every  $j$ ,  $x_j^{(n)} \rightarrow x_j$  as  $n$  goes to infinity through  $\{n_k\}$ . Furthermore, the point  $(x_1, x_2, \dots)$  is in  $C_n$  for every  $n \geq 1$ , contradicting  $C_n \downarrow \emptyset$ .

## BIBLIOGRAPHY

- [1] BACHELIER, L., "Théorie de la spéculation, Thèse, Paris, 1900," *Ann. Éc. Norm. Sup.* s. 3, **17**, 21–86 (1900).
- [2] BAILEY, N. T. J., *The Elements of Stochastic Processes*, John Wiley & Sons, Inc., New York, 1964.
- [3] BHARUCHA-REID, A. T., *Elements of the Theory of Markov Processes and Their Applications*, McGraw-Hill Book Co., Inc., New York, 1960.
- [4] BIRKHOFF, G. D., "Proof of the ergodic theorem," *Proc. Nat'l. Acad. Sci.* **17**, 656–660 (1931).
- [5] —, "Dynamical systems," *American Math. Society*, New York, 1927, reprinted 1948.
- [6] BLACKWELL, D., "A renewal theorem," *Duke Math. J.* **15**, 145–150 (1948).
- [7] —, "Extension of a renewal theorem," *Pacific J. Math.* **3**, 315–320 (1953).
- [8] —, "On a class of probability spaces," *Proc. of the 3rd Berkeley Symp. on Math. Stat. and Prob.* Vol. II., 1–6 (1956).
- [9] — and FREEDMAN, D., "The tail  $\sigma$ -field of a Markov chain and a theorem of Orey," *Ann. Math. Stat.* **35**, No. 3, 1291–1295 (1964).
- [10] BLUMENTHAL, R., GETTOOR, R., and MCKEAN, H. P., JR., "Markov processes with identical hitting distributions," *Illinois J. Math.* **6**, 402–420 (1962).
- [11] BOCHNER, S., *Harmonic Analysis and the Theory of Probability*, University of California Press, Berkeley, 1955.
- [12] BREIMAN, L., "On transient Markov chains with application to the uniqueness problem for Markov processes," *Ann. Math. Stat.* **28**, 499–503 (1957).
- [13] CAMERON, R. H. and MARTIN, W. T., "Evaluations of various Wiener integrals by use of certain Sturm-Liouville differential equations," *Bull. Amer. Math. Soc.* **51**, 73–90 (1945).
- [14] CHOW, Y. S. and ROBBINS, H., "On sums of independent random variables with infinite moments and 'fair' games," *Proc. Nat. Acad. Sci.* **47**, 330–335 (1961).
- [15] CHUNG, K. L., "Notes on Markov chains," duplicated notes, *Columbia Graduate Mathematical Statistical Society*, 1951.
- [16] —, *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag, Berlin, 1960.
- [17] —, "The general theory of Markov processes according to Doeblin," *Z. Wahr.* **2**, 230–254 (1964).
- [18] — and FUCHS, W. H. J., "On the distribution of values of sums of random variables," *Mem. Amer. Math. Soc.* No. 6 (1951).
- [19] — and KAC, M., "Remarks on fluctuations of sums of independent random variables," *Mem. Amer. Math. Soc.* No. 6 (1951).

- [20] — and ORNSTEIN, D., "On the recurrence of sums of random variables," *Bull. Amer. Math. Soc.* **68**, 30–32 (1962).
- [21] CRAMÉR, H., "On harmonic analysis in certain functional spaces," *Ark. Mat. Astr. Fys.* **28B**, No. 12 (1942).
- [22] DARLING, D. A., and SIEGERT, A. J., "The first passage problem for a continuous Markov process," *Ann. Math. Stat.* **24**, 624–639 (1953).
- [23] — and KAC, M., "On occupation times for Markov processes," *Trans. Amer. Math. Soc.* **84**, 444–458 (1957).
- [24] DINGES, H., "Ein verallgemeinertes Spiegelungsprinzip für den Prozess der Brownschen Bewegung," *Z. Wahr.* **1**, 177–196 (1962).
- [25] DOEBLIN, W., "Sur les propriétés asymptotiques de mouvement régis par certains types des chaînes simples," *Bull. Math. Soc. Roum. Sci.* **39**, No. 1, 57–115, No. 2, 3–61 (1937).
- [26] —, "Sur certains mouvements aléatoires discontinus," *Skand. Aktuarietidskr* **22**, 211–222 (1939).
- [27] —, "Sur les sommes d'un grand nombre de variables aléatoires indépendantes," *Bull. Sci. Math.* **63**, No. 1, 23–32, 35–64 (1939).
- [28] —, "Éléments d'une théorie générale des chaînes simples constante de Markoff," *Ann. Sci. École Norm. Sup.* (3), **57**, 61–111 (1940).
- [29] —, "Sur l'ensemble de puissances d'une loi de probabilité," *Studia Math.* **9**, 71–96 (1940).
- [30] DONSKE, M., "An invariance principle for certain probability limit theorems," *Mem. Amer. Math. Soc.* No. 6 (1951).
- [31] —, "Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.* **23**, 277–281, (1952).
- [32] DOOB, J. L., "Regularity properties of certain families of chance variables," *Trans. Amer. Math. Soc.* **47**, 455–486 (1940).
- [33] —, "Topics in the theory of Markov chains," *Trans. Amer. Math. Soc.* **52**, 37–64 (1942).
- [34] —, "Markov chains—denumerable case," *Trans. Amer. Math. Soc.* **58**, 455–473 (1945).
- [35] —, "Asymptotic properties of Markov transition probabilities," *Trans. Amer. Math. Soc.* **63**, 393–421 (1948).
- [36] —, "Renewal theory from the point of view of the theory of probability," *Trans. Amer. Math. Soc.* **63**, 422–438 (1948).
- [37] —, "A heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.* **20**, 393–403 (1949).
- [38] —, "Continuous parameter martingales," *Proc. 2nd Berkeley Symp. on Math. Stat. and Prob.*, 269–277 (1951).
- [39] —, *Stochastic Processes*, John Wiley & Sons, Inc., New York, 1953.
- [40] DUBINS, L. and SAVAGE, J. L., *How to Gamble If You Must*, McGraw-Hill Book Co., Inc., New York, 1965.
- [41] DUNFORD, N. and SCHWARTZ, J. T., *Linear Operators, Part I*, Interscience Publishers, Inc., New York, 1958.
- [42] DVORETSKI, A., ERDÖS, P., and KAKUTANI, S., "Nonincrease everywhere of the Brownian motion process," *Proc. 4th Berkeley Symp. on Math. Stat. and Prob.* Vol. II, 103–116 (1961).

- [43] DYNKIN, E. B., "Continuous one-dimensional Markov processes," *Dokl. Akad. Nauk. SSSR* **105**, 405-408 (1955).
- [44] —, *Markov processes*, Vols. I, II, Academic Press, Inc., New York, 1965.
- [45] EINSTEIN, A., "On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat," *Ann. d. Physik* **17** (1905). [In *Investigations of the Theory of the Brownian Movement*, Edited by R. Fürth, Dover Publications, Inc., New York, 1956.]
- [46] ERDÖS, P., "On the law of the iterated logarithm," *Ann. of Math.* **43**, 419-436 (1942).
- [47] — and KAC, M., "On certain limit theorems of the theory of probability," *Bull. Amer. Math. Soc.* **52**, 292-302 (1946).
- [48] —, "On the number of positive sums of independent random variables," *Bull. Amer. Math. Soc.* **53**, 1011-1020 (1947).
- [49] —, FELLER, W., and POLLARD, H., "A theorem on power series," *Bull. Amer. Math. Soc.* **55**, 201-204, 1949.
- [50] FELLER, W., "Zur Theorie der stochastischen Prozesse," *Math. Ann.* **113**, 113-160 (1936).
- [51] —, "On the Kolmogorov-P. Lévy formula for infinitely divisible distribution functions," *Proc. Yugoslav Acad. Sci.* **82**, 95-113 (1937).
- [52] —, "On the integro-differential equations of purely discontinuous Markov processes," *Trans. Am. Math. Soc.* **48**, 488-515 (1940).
- [53] —, "The general form of the so-called law of the iterated logarithm," *Trans. Amer. Math. Soc.* **54**, 373-402 (1943).
- [54] —, "A limit theorem for random variables with infinite moments, *Amer. J. Math.* **68**, 257-262 (1946).
- [55] —, "Fluctuation theory of recurrent events," *Trans. Amer. Math. Soc.* **67**, 98-119 (1949).
- [56] —, "Diffusion processes in genetics," *Proc. 2nd Berkeley Symp. on Math. Stat. and Prob.* 227-246 (1951).
- [57] —, "The parabolic differential equations and the associated semi-groups of transformations, *Ann. of Math.* **55**, 468-519 (1952).
- [58] —, "Diffusion processes in one dimension," *Trans. Amer. Math. Soc.* **77**, 1-31 (1954).
- [59] —, *An Introduction to Probability Theory and Its Applications*, Vol. I, 2nd Ed., Vol. II, John Wiley & Sons, Inc., New York, 1957, 1966.
- [60] — and OREY, S., "A renewal theorem," *J. Math. and Mech.* **10**, 619-624 (1961).
- [61] GARSIA, A., "A simple proof of Eberhard Hopf's maximal ergodic theorem," *J. Math. and Mech.* **14**, 381-382 (1965).
- [62] GNEDENKO, B. V. and KOLMOGOROV, A. N., *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1954.
- [63] HADLEY, G., *Linear Algebra*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1961.
- [64] HALMOS, P. R., *Measure Theory*, D. Van Nostrand Co., Inc., Princeton, N.J., 1950.
- [65] —, "Lectures on ergodic theory," *Mathematical Society of Japan*, No. 3, 1956.

- [66] HARRIS, T. E., "The existence of stationary measures for certain Markov processes," *Proc. of the 3rd Berkeley Symp. on Math. Stat. and Prob.* Vol. II, 113-124 (1956).
- [67] —, *The Theory of Branching Processes*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.
- [68] —, "First passage and recurrence distributions," *Trans. Amer. Math. Soc.* **73**, 471-486 (1952).
- [69] — and ROBBINS, H., "Ergodic theory of Markov chains admitting an infinite invariant measure," *Proc. of the National Academy of Sciences* **39**, No. 8, 862-864 (1953).
- [70] HERGLOTZ, G., "Über Potenzreihen mit positivem reellen Teil im Einheitskreis," *Ber. Verh. Kgl. Sachs. Ges. Leipzig, Math.-Phys. Kl.* **63**, 501-511 (1911).
- [71] HEWITT, E. and SAVAGE, L. J., "Symmetric measures on Cartesian products," *Trans Amer. Math. Soc.* **80**, 470-501 (1955).
- [72] HOBSON, E. W., *The Theory of Functions of a Real Variable*, 3rd Ed., Vol. I, Cambridge University Press, Cambridge, 1927 (Reprinted by Dover Publications, Inc., New York, 1957).
- [73] HOPF, E., *Ergoden Theorie, Ergebnisse der Math.* Vol. 2, J. Springer, Berlin, 1937 (Reprinted by Chelsea Publishing Co., New York, 1948).
- [74] HUNT, G., "Some theorems concerning Brownian motion," *Trans. Amer. Math. Soc.* **81**, 294-319 (1956).
- [75] ITO, K., "On stochastic processes (I) (Infinitely divisible laws of probability)," *Japan J. Math.* **18**, 261-301 (1942).
- [76] — and McKEAN, H. P. JR., *Diffusion Processes and Their Sample Paths*, Academic Press, New York, 1965.
- [77] JESSEN, B. and SPARRE ANDERSEN, E., "On the introduction of measures in infinite product sets," *Danske Vid. Selsk. Mat.-Fys. Medd.* **25**, No. 4 (1948).
- [78] — and WINTNER, A., "Distribution functions and the Riemann Zeta function," *Trans. Amer. Math. Soc.* **38**, 48-88 (1935).
- [79] KAC, M., "On a characterization of the normal distribution," *Amer. J. of Math.* **61**, 726-728 (1939).
- [80] —, "Random walk and the theory of Brownian motion," *Amer. Math. Monthly* **54**, 369-391 (1949) (Reprinted in [139]).
- [81] —, "On the notion of recurrence in discrete stochastic processes," *Bull. Amer. Math. Soc.* **53**, 1002-1010 (1947).
- [82] —, "On some connections between probability theory and differential and integral equations," *Proc. 2nd Berkeley Symp. on Math. Stat. and Prob.* 189-215 (1951).
- [83] —, "Statistical independence in probability, analysis, and number theory," *Carus Mathematical Monograph*, No. 12, The Mathematical Association of America, 1959.
- [84] KALLIANPUR, G. and ROBBINS, H., "The sequence of sums of independent random variables," *Duke Math. J.* **21**, 285-307 (1954).
- [85] KARAMATA, J., "Neuer Beweis und Verallgemeinerung der Tauberschen Sätze, welche die Laplaceschen Stieltjesschen Transformationen betreffen," *J. für die reine und angewandte Math.* **164**, 27-40 (1931).

- [86] KARLIN, S., *A First Course in Stochastic Processes*, Academic Press, New York, 1966.
- [87] — and MCGREGOR, J. L., "Representation of a class of stochastic processes," *Proc. Nat'l. Acad. Sci.* **41**, 387–391 (1955).
- [88] KHINTCHINE, A., "Über einen Satz der Wahrscheinlichkeitsrechnung," *Fundamenta Math.* **6**, 9–20 (1924).
- [89] —, "Dédution nouvelle d'une formula de M. Paul Lévy," *Bull. Univ. d'Etat Moskou.*, Sér Internat. Sect. A., 1, No. 1, 1–5 (1937).
- [90] —, *Mathematical Foundations of Statistical Mechanics*, Dover Publications, Inc., New York, 1949.
- [91] — and KOLMOGOROV, A., "Über konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden," *Rec. Math (Mat. Sbornik)* **32**, 668–677 (1925).
- [92] KOLMOGOROV, A., "Sur la loi forte des grandes nombres," *C. R. Acad. Sci. Paris*, **191**, 910–912 (1930).
- [93] —, "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung," *Math. Ann.* **104**, 415–458 (1931).
- [94] —, "Sulla forma generale di un processo stocastico omogeneo," *Atti Acad. naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (6), **15**, 805–808, 866–869 (1932).
- [95] —, "Anfangsgründe der Theorie der Markoffschen ketten mit unendlich vielen möglichen Zuständen," *Rec. Math. Moscov (Mat. Sbornik)*, **1** (43), 607–610 (1936).
- [96] —, "Interpolation and extrapolation of stationary random sequences" (Russian), *Izv. Akad. Nauk. SSSR, Ser. Mat.* **5**, 3–114 (1941).
- [97] —, "Stationary sequences in Hilbert space" (Russian), *Bull. Math. Univ. Moscov* **2**, No. 6 (1941).
- [98] —, *Foundations of Probability* (translation), Chelsea Publishing Co., New York, 1950.
- [99] LE CAM, L., Mimeographed notes, Statistics Department, Univ. of Calif., Berkeley.
- [100] LÉVY, P., "Théorie des erreurs. La loi de Gauss et les lois exceptionnelles," *Bull. Soc. Math.* **52**, 49–85 (1924).
- [101] —, "Sur les séries dont les termes sont des variables éventuelles indépendantes," *Studia Math.* **3**, 119–155 (1931).
- [102] —, "Sur les intégrales dont les éléments sont des variables aléatoires indépendantes," *Annali R. Scuola. Sup. Pisa* (2), **3**, 336–337 (1934) and **4**, 217–218 (1935).
- [103] —, *Théorie de l'Addition des Variables Aléatoires*, Bautier-Villars, Paris, 1937.
- [104] —, "Sur certains processus stochastiques homogènes," *Comp. Math.* **7**, 283–339 (1939).
- [105] —, *Processus Stochastiques et Mouvement Brownian*, Gauthier-Villars, Paris, 1948.
- [106] LINDBERG, Y. W., "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitrechnung," *Math. Z.* **15**, 211–225 (1922).
- [107] LOÈVE, M., Appendix to [105].
- [108] —, *Probability Theory*, 3rd Ed., D. Van Nostrand Co., Inc., Princeton, N.J., 1963.

- [109] LYAPUNOV, A. M., "Nouvelle forme du théorème sur la limite de probabilités, *Mem. Acad. Sci. St. Petersburg* (8), 12, No. 5, 1-24 (1901).
- [110] MARKOV, A. A., "Extension of the law of large numbers to dependent events" (Russian), *Bull. Soc. Phys. Math. Kazan* (2) 15, 155-156 (1906).
- [111] MARUYAMA, G., "The harmonic analysis of stationary stochastic processes," *Mem. Fac. Sci. Kyusyu Univ.* A4, 45-106 (1949).
- [—a] McSHANE, E. J., *Integration*, Princeton University Press, Princeton, N.J., 1944.
- [112] MISES, R. v., *Probability, Statistics, and Truth*, William Hodge and Co., 2nd Ed., London, 1957.
- [113] NEVEU, J., *Mathematical Foundations of the Calculus of Probability*, Holden-Day, Inc., San Francisco, 1965.
- [114] OREY, S., "An ergodic theorem for Markov chains," *Z. Wahr.* 1, 174-176 (1962).
- [115] PALEY, R., WIENER, N., and ZYGMUND, A., "Note on random functions," *Math. Z.*, 647-668 (1933).
- [116] POLYA, G., "Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt in Strassennetz," *Math. Ann.* 89, 149-160 (1921).
- [117] POSPIŠIL, B., "Sur un problème de M. M. S. Bernstein et A. Kolmogoroff," *Casopis Pest. Mat. Fys.* 65, 64-76 (1935-36).
- [118] PROKHOROV, Y. V., "Convergence of random processes and limit theorems in probability theory," *Teor. Veroyatnost. i. Primenen* 1, 177-238 (1956).
- [119] RYLL-NARDEWSKI, C., "Remarks on processes of calls," *Proc. of the 4th Berkeley Symp. on Math. Stat. and Prob.* Vol. II, 455-465 (1961).
- [120] SAATY, T. L., *Elements of Queuing Theory, with Applications*, McGraw-Hill Book Co., Inc., New York, 1961.
- [121] SAKS, S., "Theory of the integral," *Monografie Matematyczne*, Tom VII, Warsaw-Lwow (1937).
- [122] SHEPP, L. A., "A local limit theorem," *Ann. Math. Stat.* 35, 419-423 (1964).
- [123] SHOHAT, J. A. and TAMARKIN, J. D., "The problem of moments," *Math. Surveys*, No. 1, *Amer. Math. Soc.*, New York (1943).
- [—a] SIERPINSKI, W., *General Topology*, University of Toronto Press, Toronto, 1952.
- [124] SKOROKHOD, A. V., "Limit theorems for stochastic processes," *Teor. Veroyatnost. i. Primenen* 1, 289-319 (1956).
- [125] —, "Limit theorems for stochastic processes with independent increments," *Teor. Veroyatnost. i. Primenen* 2, 145-177 (1957).
- [126] —, *Studies in the Theory of Random Processes*, Kiev University, 1961; Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965 (translation).
- [127] SMITH, W. L., "Renewal theory and its ramifications," *J. Roy. Statist. Soc.* (3), 20, 243-302 (1958).
- [128] SPARRE ANDERSEN, E., "On the fluctuations of sums of random variables," *Math. Scand.* 1, 163-285 (1953) and 2, 195-223 (1954).
- [129] SPITZER, F., "A combinatorial lemma and its applications to probability theory," *Trans. Amer. Math. Soc.* 82, 323-339 (1956).
- [130] —, *Principles of random walk*, D. Van Nostrand Co., Inc., Princeton, 1964.
- [131] STONE, C. J., "Limit theorems for random walks, birth and death processes, and diffusion processes," *Illinois J. Math.* 7, 638-660 (1963).

- [132] —, "On characteristic functions and renewal theory," *Trans. Amer. Math. Soc.* **120**, 327–342 (1965).
- [133] —, "A local limit theorem for multidimensional distribution functions," *Ann. Math. Stat.* **36**, 546–551 (1965).
- [134] STRASSEN, V., "An invariance principle for the law of the iterated logarithm," *Z. Wahr.* **3**, 211–226 (1964).
- [135] —, "A converse to the law of the iterated logarithm," *Z. Wahr.* **4**, 265–268 (1965).
- [136] TROTTER, H. F., "A property of Brownian motion paths," *Illinois J. Math.* **2**, 425–433 (1958).
- [137] VILLE, J., *Etude Critique de la Notion de Collectif*, Gauthier-Villars, Paris, 1939.
- [138] VOLKONSKIĬ, V. A., "Random substitution of time in strong Markov processes," *Teor. Veroyatnost. i. Primenen* **3**, 332–350 (1958).
- [139] WAX, N. (editor), *Selected Papers on Noise and Stochastic Processes*, Dover Publications, Inc., New York, 1954.
- [140] WIDDER, D. V., *The Laplace Transform*, Princeton Univ. Press, Princeton, 1941.
- [141] WIENER, N., "Differential space," *J. Math. Phys.* **2**, 131–174 (1923).
- [142] —, "Un problème de probabilités énombrables," *Bull. Soc. Math. de France* **52**, 569–578 (1924).
- [143] —, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, MIT Press and John Wiley & Sons, Inc., New York, 1950 (reprinted from a publication restricted for security reasons in 1942).
- [144] YAGLOM, A. M., *An Introduction to the Theory of Stationary Random Functions*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1962.



## INDEX

- Absorbing boundary, for Brownian motion, 353
- for a Feller process, 369–370
- Absorption probabilities, for Markov chains, 155–156
- for fair coin-tossing, 156
- Accessible boundary point, 366
- Andersen, E., 103, 297
- Arc sine law, 272, 282, 283, 297
- Asymptotic behavior of Markov transition probabilities, general state space, 137
- countable state space, 150
- Orey's theorem, 151
- continuous time processes, 344
- Asymptotic stationarity, for Markov chains, 133–135
- for renewal processes, 147
- Bachelier, L., 271
- Backwards equations, hold for pure jump processes, 331–332, 335
- for Feller processes, 385
- Bailey, N., 346
- Berry-Eseen bound, 184
- Bharucha-Reid, A., 158, 346
- Birkhoff, G., 104, 109, 127, 128
- Birth and death processes, discrete time, 147–148
- continuous parameter, 337–339
- Blackwell, D., 35, 62, 81, 101, 151, 231
- Bochner, S., 317
- Bochner's theorem on characteristic functions, 174
- Borel, E., 65
- Borel-Cantelli lemma, proof, 41–42
- applied to recurrence in coin-tossing, 42–44
- generalized to arbitrary processes, 96–97
- Borel field, in sequence space, 12
- finite-dimensional, 22
- in function space, 251
- Borel paradox, 73
- Boundaries, for a Brownian motion, 352–355
- for Feller processes, 365–370
- Boundary conditions, for a Markov process on the integers, 339–340
- for a Brownian motion, 352–355
- Bounded pointwise convergence, 340–341
- Branching processes, 148–149
- Brown, R., 271
- Brownian motion, defining assumptions, 248
- normally distributed, 249–250
- distribution of, 250
- as limit in distribution of random walks, 251
- joint measurability, 257
- normalized, 257
- continuous sample paths, 257–259
- distribution of the maximum, 258, 287–288
- defined by an infinite expansion, 259–261
- nondifferentiable, 261–262
- unbounded variation, 262
- variation of paths, 262–263
- law of the iterated logarithm, 263–265

- behavior for large time, 265–267
  - time inversion, 266
  - set of zeros, 267–268
  - stopping times, 268–270
  - strong Markov property, 268–270
  - first-exit distributions, 273–275, 287–290
  - used to represent sums of random variables, 276–278
  - as uniform limit of normed sums, 278–281
  - as limit in invariance theorem, 281–283
  - transformations, 287
  - distribution of first hitting times, 287–288
  - with stable subordinator, 317–318
  - infinitesimal operator, 325–326
  - backwards and forwards equation, 327–328
  - resolvent, 344
  - exact model, 347–350
  - boundaries, 352–355
  - scale, 361
  - speed measure, 365
  - used to construct Feller processes, 370–375
  - local time, 373–374
- Cameron, R., 297
- Central limit problem, conditions for convergence of distributions, 198–199
- convergence to the normal, 198–199
- convergence to stable laws, 199–200, 207–212
- Central limit problem generalized, 195–196
- Central limit theorem, for coin-tossing, 7–10
- for identically distributed random variables with finite second moment, 167–168
- nonidentically distributed case, 186–187
- in terms of distribution of maximum, 190, 198–199
- for multidimensional variables, 237–238
- Chapman-Kolmogorov equations, 320–321, 322
- Characteristic functions, defined, 170
- properties, 170–171
- and independent random variables, 175–177
- inversion formulas, 177–179
- expansion, 180
- logarithms of, 180–181
- normal distribution, 188
- Poisson distribution, 189
- infinitely divisible laws, 191–195
- stable laws, 200–203, 204–207
- of multidimensional distributions, 235–236
- of multivariate normal distribution, 237, 239
- of processes with independent increments, 304–305
- Characteristic operator, defined, 375–376
- as second order differential operator, 377–379
- minimum principle for, 380
- uniqueness of solutions, 380–382
- for a diffusion, 386
- Chebyshev inequality for coin-tossing, 4
- Chow, Y., 65
- Chung, K. L., 65, 145, 157, 158, 231, 232, 332, 339, 346
- Coin-tossing, probability of an equalization, 2
- weak law of large numbers, 3
- strong law of large numbers, 11–13
- analytic model, 15
- equalizations, 62
- probability of no equalizations in biased coin-tossing, 141
- absorption probabilities, 155
- limit distribution of returns to equilibrium, 214
- Communicating states, 141

- Compactness of the set of distribution functions, 160–163
  - multidimensional, 234–235
- Conditional expectation, given one random variable, 70–71
  - given a  $\sigma$ -field, 73–76
- Conditional probabilities, for a Markov process, 321–322
- Conditional probability, of one event
  - given another, 67
  - given the value of a random variable, 68–69
- Continuity of Brownian motion sample paths, 257–259
- Continuity in probability, 299
- Continuity theorem, for characteristic functions, 171–172
  - in terms of Laplace transforms, 183
  - in terms of moment-generating functions, 183
  - for multidimensional characteristic functions, 236
- Continuous parameter processes, 248, 251–253
- Continuous versions of processes, 254–257
- Convergence definitions, almost surely,  $\xrightarrow{\text{a.s.}}$ , 33
  - in mean  $\xrightarrow{r}$ , 33
  - in probability,  $\xrightarrow{P}$ , 33
  - in distribution  $\xrightarrow{D}$ , 159, 233
  - uniformly on compact sets  $\xrightarrow{uc}$ , 173
  - weak  $\xrightarrow{w}$ , 217
  - boundedly pointwise  $\xrightarrow{bp}$ , 341
- Convergence in distribution, definition, 159
  - behavior of integrals, 163–164
  - separating classes of functions, 165–166
  - convergence of expectations, 167
  - determined by characteristic functions, 170–173
  - and convergence of moments, 181–182
  - of random vectors, 233–234
  - of processes, 281, 293
- Convergence of integrals, 163–164
- Convergence of sequences of conditional expectations, 92–94
- Convergence of types theorem, 174–175
- Coordinate representation process, 22
- Covariance function, 241, 243
- Covariance matrix, 237, 239
- Cramér, H., 247
- Darling, D., 232, 383
- Decomposition of processes with stationary independent increments, 310–315
- Degenerate distribution, 160
- Diffusions, backwards equations, 352
  - infinitesimal operator, 352
  - construction, 370–372
  - defined, 385–386
  - scale and speed measure, 386–387
  - characterization of, 387–389
  - as Feller processes, 387–389
- Dinges, H., 297
- Distribution functions, determine the distribution of a process, 26
- Distribution of pure type, 49
- Distribution of a sum of two independent random variables, 39
- Doob, J. L., 81, 101, 102, 103, 128, 137, 157, 220, 247, 290, 297, 306, 318, 335, 346, 390
- Dubins, L., 34, 103
- Dunford, N., 128
- Dvoretzki, A., 261
- Dynkin, E., 346, 390
- Dynkin's identity, 376
- Ehrenfest urn scheme, 149
- Einstein, A., 271

- Entrance boundary points, 366
- Equidistribution of sums of independent random variables, 58–61
- Erdős, P., 231, 261, 296, 297
- Ergodic theorem, proof, 113–115
  - applied to orthogonality of measures, 116
  - converse, 116
  - convergence in first mean, 117
  - for stationary processes, 118
  - applied to the range of sums, 120–122
- Ergodicity, of measure-preserving transformation, 109–110, 112–113
  - of stationary processes, 119
  - of process of independent, identically distributed random variables, 120
  - of process of recurrence times for a stationary ergodic process, 124–125
- Markov chains, 136–137
  - of Gaussian stationary processes, 246
- Exit boundary points, 368–369
- Expectation, defined, 31
  - of products of independent random variables, 39
- Explosions, in a Markov process, 336–338
- Exponent function, 305
- Extended random variable, 30
- Extension of probabilities, in sequence space, 23–25
  - from distribution functions in discrete time, 28
  - for Markov chains, 130–131
  - on function spaces, 251–253
  - by continuity of paths, 254–257
  - for Markov processes, 321
- Extensions to smooth versions, 298–300
- Feller-Orey renewal lemma, 221–222
- Feller process, defined, 356–358
  - stability of transition probabilities, 356–357, 358
  - regularity, 358
  - scale, 358–362
  - first-exit distributions, 359–361, 383–384
  - speed measure, 362–365
  - boundary classification, 365–370
  - construction, 370–375
  - local time, 373
  - characteristic operator, 375–379
  - uniqueness, 379–383
  - densities of transition probabilities, 384
  - resolvent, 384
  - backwards equation, 385
  - stationary initial distribution, 385
  - (See Diffusions)
- Feller, W., 18, 62, 63, 64, 150, 158, 211, 215, 221, 228, 231, 232, 297, 346, 390
- Field of events depending on a finite number of outcomes, 30
- Finetti, B. de, 318
- First-exit distributions, for Brownian motion, 273–275, 287–290
  - from linear boundaries, 289–290
  - for Feller processes, 359–361, 383–384
- First-exit time, for Brownian motion, 273
  - finite expectation, 359
  - and speed measure, 363
- Fourier inversion formulas, 177–179
- Freedman, D., 151, 271
- Fuchs, W., 65, 145
- Functionals on Brownian motion, 282
- Gambling systems, 82–83, 101–102
- Garsia, A., 114
- Gaussian stationary processes,
  - definition, 241, 242
  - spectral representation, 242–246
- ergodicity, 246
- prediction problem, 246–247
- Ornstein-Uhlenbeck process, 347–350
- Gibbs, W., 109
- Gnedenko, B., 215, 232
- Green's function, 363, 365, 367–368
- Halmos, P., 35, 91, 128
- Harris, T., 65, 148, 149, 157, 346
- Helly-Bray theorem, 160–161
  - for multidimensional distributions, 234–235

- Herglotz lemma, 242  
 Hewitt, E., 63  
 Hewitt-Savage zero-one law, proof, 63  
 Hopf, E., 128  
 Hunt, G., 271
- Independence, definitions, 36  
 Independent events, 41, 44  
 Independent increments, 303  
 Independent  $\sigma$ -fields, 36  
 Infinite sums of independent random variables, convergence almost surely, 45–48  
     finiteness of sums of moments, 47, 48  
     distribution of, 49–51  
     converge or diverge badly, 97  
     convergence in distribution, 176  
     convergence of characteristic functions of, 177  
 Infinitely divisible distribution of processes with independent increments, 303–305  
 Infinitely divisible laws, defined, 191  
     as limits of distributions of sums, 191  
     characteristic function, 194  
     as limits of distributions of nonidentically distributed summands, 195–196  
     necessary and sufficient conditions for convergence to, 198  
 Infinitesimal conditions, for a pure jump process, 332–333  
 Infinitesimal operator, domain, general case, 341  
     for a diffusion, 352  
 Instantaneously reflecting boundary points, 369  
 Invariance principle, applied to central limit theorem, 167–169  
     for sums of independent random variables, 272–273, 281–283  
     applied to Kolmogorov-Smirnov statistics, 283–287  
     applied to the law of the iterated logarithm, 291–292  
     for general processes, 293–296  
 Ito, K., 271, 318, 373, 390
- Jensen's inequality, 80  
 Jessen, B., 44, 103  
 Joint density, 71  
 Joint normal distribution, 237, 238–240  
 Jump processes, 310–315  
 Jump times, for a Markov process, 328–332  
 Jumps of a process with independent increments, 310–315
- Kac, M., 18, 48, 128, 150, 232, 296, 297, 351  
 Kakutani, S., 261  
 Kallianpur, G., 232  
 Karamata, J., 228  
 Karlin, S., 148, 158  
 Kesten, H., 120  
 Khintchine, A., 65, 128, 215, 263, 271  
 Kolmogorov, A. N., 34, 40, 48, 65, 73, 81, 158, 215, 231, 232, 246, 247, 346  
 Kolmogorov inequality, 65  
 Kolmogorov zero-one law, 40  
 Kolmogorov-Smirnov statistics, 283–287, 290  
 Kronecker's lemma, 51
- Langevin equation, 351  
 Lattice distribution, 54, 174, 225  
 Law of the iterated logarithm, for sums of independent random variables, 64, 291–292  
     for Brownian motion, 263–265, 266  
 Law of pure types, 49  
 Law of a random variable, 159  
 Lebesgue bounded convergence theorem, 33  
 Le Cam, L., 184  
 Lévy, P., 51, 66, 103, 215, 271, 297, 318  
 Lindeberg, J., 184  
 Local central limit theorem, 224–227  
 Local time, 373–375  
 Loève, M., 35, 48, 66, 184, 247  
 Lyapunov, A., 184, 215
- Markov, A., 157  
 Markov chains, definitions, 129–131  
     construction of, 130–131

- Markov time, 131
- strong Markov property, 131–132
- asymptotic stationarity, 133–135
- stationary initial distributions, 134, 136–137, 143–145
- closed sets, 135
- indecomposability, 135–137
- ergodicity, 136–137
- periodic motion, 137, 140
- backward difference equations, 153–156
- Markov chains with a countable state space, renewal process of a state, 138
- returns to a state, 138–140
- classification of states, 139–140
- recurrence times, 139–140
- group properties of states, 141–143
- cyclically moving subsets, 142–143
- stationary initial distributions, 143–145
- renewal chain, 146–147
- birth and death processes, 147, 148
- branching processes, 148–149
- Ehrenfest urn scheme, 149
- asymptotic behavior of transition probabilities, 150–153
- tail  $\sigma$ -field, 152–153
- random variables invariant under shifts, 154
- absorption, probabilities, 154–155
- time until, 155–156
- Markov processes, defined, 319–320
- transition probabilities, 319–323
- strong Markov property, 323
- infinitesimal operator, 325–326
- backwards and forwards equations, 326–328
- with pure jump sample paths defined, 328
- uniqueness, general case, 340–343
- resolvent, 340–344
- domain of the infinitesimal operator, 341
- stationary initial distributions, 344, 346
- Ornstein-Uhlenbeck processes, 347–350
- locally Brownian processes, 351–352
- (See Diffusions; Feller processes; Pure jump Markov processes)
- Markov processes moving on the integers, returns from infinity, 339–340
- backwards and forwards equations, 340
- asymptotic behavior of transition probabilities, 344–345
- Markov times, for Brownian motion, 268–270
- defined, 323
- Martin, R., 297
- Martingales and submartingales, definitions, 83–84
- optional sampling theorem, 84–89
- inequalities, 88–89
- strong convergence theorem, 89–90
- upcrossing lemma, 91
- as conditional expectations, 92–94
- optional stopping, 95–96
- applied to the generalized Borel-Cantelli lemma, 96–97
- stopping rules, 98–100
- Wald's identity, 100
- applied to gambling systems, 101–102
- continuous parameter, 274, 300–303
- continuity properties of sample paths, 300–301
- optional stopping for continuous parameter martingales, 302–303
- applied to processes with independent increments, 306–307
- Mass-preserving sets of distributions, 162
- Mass-preserving sets of multi-dimensional distributions, 234
- Maximal ergodic theorem, 114
- Maximum of a Brownian motion, 258, 287–288
- Maximum of independent random variables, 189–190
- McKean, H., 271, 373 390
- Measure-preserving transformations, definition, 106
- invariant sets, 108
- invariant  $\sigma$ -field, 108
- ergodicity, 109–110, 112–113

- invariant random variables, 112–113
  - ergodic theorem, 113–115
- Method of moments, 181–182
- Minimal solution, for Markov process, 336
- Mises, R. v., 34
- Moment-generating function, 183
- Moment problem, 182
- Moments, applied to distribution of occupation times, 229–231
- Multivariate normal distribution, 237, 238–240
- Natural boundary points, 366
- Natural scale, 358–362
- Neveu, J., 18, 35
- Nondifferentiability of Brownian paths, 261–262
- Normal distribution, 9, 170, 185–186, 188
  - domain of attraction, 214–215
  - invariance under orthogonal rotations, 350
- Normalized Brownian motion, 257
- Null-recurrent states, 140
- Occupation times, for sums of lattice random variables, 229
  - for sums of nonlattice random variables, 229–231
- Occurrence infinitely often, 40
- Optional sampling, 85
- Optional sampling theorem, 84–89
- Optional skipping, 49
- Order statistics, 285–286
- Orey, S., 151, 221, 231
- Ornstein, D., 65
- Ornstein-Uhlenbeck process, 347–350
- Paley, R., 271
- Paths with jump discontinuities, 298–300
- Periodic states, 140
- Point processes, stationary, 125–127
- Poisson, convergence, 188–190
  - distribution, 189
  - process, 308–310
  - processes with random jumps, 310–312
- Pollard, H., 231
- Polya, G., 65, 145
- Pospíšil, B., 346
- Probabilities on function spaces, 251–253, 254–257
- Probability space, 14, 19
- Process, continuous parameter, defined, 248, 251–253
- Process, discrete time, definition, 19
  - measurability, 21
  - distribution of, 22
- Processes with stationary, independent increments, defined, 303
  - infinitely divisible distributions, 303–305
  - path properties, 306–307
  - Poisson process, 308–310
  - jump processes, 310–312
  - limits of jump processes, 312–315
  - as sums of independent jump processes, 312–315
  - zero-one law for path properties, 315
  - as Markov processes, 324
  - infinitesimal operator, 325
- Prokhorov, Y., 297
- Pure jump function, 314
- Pure jump Markov processes, defined, 328
  - strong Markov property, 328
  - first jump distributions, 328–330
  - backwards equations, 331–332, 335
  - infinitesimal conditions, 332–333
  - construction, 332–336
  - space and time structure, 333–334
  - uniqueness, 334–336, 339–340
  - existence, 336–339
  - explosions, 336–339
  - birth and death processes, 337–339
  - conditions for no accumulation of jumps, 337–339
  - nonuniqueness and boundary conditions, 339–340
  - minimal solution, 336
- Random boundaries for Brownian motion, 276–278
- Random signs problem, 41, 45, 47

- Random variables, definition, 19
  - sufficient conditions for
    - measurability, 30
  - distribution of, 31
  - expectation of, 31
  - identically distributed, 31
  - strong convergence defined, 33
  - Cauchy convergence, 33, 44
  - independence, 36, 37, 38
  - uniform integrability, 91, 94
- Random vectors, definition, 20
  - distribution of, 21
  - function of, 30
- Random walk, 132, 145
- Range of sums of independent variables, 120–122
- Recurrence, in coin-tossing, 42
  - in sums of independent random variables, 53–58
  - in Markov chains, 138–141
- Recurrence of sums of independent variables, determined by characteristic function, 179
- Recurrence times, for sums of independent random variables, 60
  - for a stationary process, 122–125
  - for a Markov state, 139–140
  - for the Ehrenfest urn scheme, 149–150
  - for sums of lattice random variables, 229
  - for a Markov process, 344–345
- Recurrent states, 139
- Reflecting boundary, for Brownian motion, 353–355
  - for a Feller process, 369–370
- Regular boundary points, 368–369
- Regular conditional distribution,
  - definition, 77
  - existence, 78–79
- Regular conditional probability, 77, 79
- Regular Feller processes, 358
- Regular transition probabilities, 321
- Renewal process, of a Markov state, 138
  - as a Markov chain, 146–147
- Renewal theorem, discrete time, 150
  - general case, 218–224
  - applied to Markov processes, 344–345
- Resolvent, of a Markov process, 340–344
  - and uniqueness of Markov processes, 343
  - for Feller processes, 382, 384
- Riemann-Lebesgue lemma, 216–217
- Robbins, H., 65, 232
- Runs in coin-tossing, 140
- Ryll-Nardzewski, C., 128
- Saaty, T., 346
- Sample path properties, Brownian motion (*See* Brownian motion)
  - general conditions to have only jump discontinuities, 298–300
  - of continuous parameter martingales, 300–301
  - of processes with stationary, independent increments, 306–307, 312–315, 315–316
  - of Poisson processes, 308–310
- Savage, J., 34, 63, 103
- Scale, of a Feller process, 358–362
  - determined by characteristic operator, 379
  - for a diffusion, 386
- Schwartz, J., 128
- Second-order stationary processes, 247
- Separating classes of functions, properties, 165–166
  - complex exponentials, 170
  - the polynomials do not separate, 181
  - real exponentials, 183
  - for multidimensional distributions, 235
- Shepp, L., 232
- Shift transformation, 107, 118
- Shohat, J., 184
- Siebert, A. J., 383
- Skorokhod, A., 293, 297
- Skorokhod's lemma, 45
- Slowly reflecting boundary points, 369–370



- Smith, W., 231
- Spectral distribution function, 242–243
- Spectral integral, 243
- Spectral representation theorem, 244–246
- Speed measure, 362–365  
   on boundary points, 367–369  
   absolutely continuous case, 370–372, 386–389  
   determined by characteristic operator, 379  
   for a diffusion, 387
- Spitzer, F., 66, 120, 232, 297
- Stable laws, defined, 199  
   as limits in distribution of normed sums, 199–200  
   characteristic functions of, 200–203, 204–207  
   domains of attraction, 207–212
- Stable processes, 316–318
- Stable subordinator, 317–318
- Stable transition probabilities, for Markov chains, 163  
   for Feller processes, 356–357, 358
- Stationary initial distribution, for Markov chains, 134, 136–137, 143–145  
   for Markov processes, 346  
   for Feller processes, 385
- Stationary point processes, 125–127
- Stationary processes, definition, 104  
   transformation of, 105  
   ergodic theorem, 118  
   ergodicity, 119  
   Gaussian, 241–247  
   second order, 247
- Stationary transition probabilities, 129, 322–323
- Stirling's approximation, 2
- Stochastic process, continuous  
   parameter, defined, 248, 251–253  
   distribution of, 252  
   with continuous sample paths, 255–257  
   having only jump discontinuities, 298–300
- Stone, C., 231, 390
- Stopping times, for sums of independent random variables, 59, 62  
   for martingales, 95–97  
   for Markov chain, 131  
   for Brownian motion, 268–270  
   for Brownian motion with random boundaries, 276–278  
   for continuous parameter martingales, 301–303  
   for Markov processes, 323
- Strassen, V., 66, 291, 297
- Strong law of large numbers, for coin-tossing, 11–13  
   for nonidentically distributed, independent random variables, 51–52  
   for identically distributed, independent random variables, 52–53  
   for Brownian motion, 265–266
- Strong Markov property, for discrete Markov chains, 131–132  
   for Brownian motion, 268–270  
   for general Markov processes, 323  
   for pure jump processes, 328  
   for Feller processes, 357
- Sums of independent random variables, convergence of, 45–49  
   law of large numbers for, 51–53  
   recurrence properties, 54–58  
   equidistribution of, 58–61  
   stopping times, 59  
   tail  $\sigma$ -field, 64  
   are martingales, 83  
   limit theorem for their range, 120–122  
   renewal theorem, 218–224  
   invariance principle, 272–273, 281–283  
   distribution of the maximum, 273  
   representation by Brownian motion, 276–278  
   converge strongly to Brownian motion paths, 278–281  
   law of the iterated logarithm, 291–292

- Tail events, defined, 40
- Tail  $\sigma$ -field, for independent random variables, 40
  - for sums of independent random variables, 64
  - condition for the zero-one property, 95
  - for Markov chains with a countable state space, 152–153
- Tamarkin, J., 184
- Tauberian theorem, 228
  - applied to recurrence times, 228–229
  - applied to occupation times, 229–231
- Transformations of Brownian motion, 287
- Transient states, 139
- Transition probabilities, stationary, 129, 322
  - for continuous parameter Markov processes, 320–321
  - regular, 321
  - standard, 323
  - backwards and forwards equations, 326–328
- Trotter, H., 373
- Uniform integrability, 91, 94
- Upcrossing lemma, 91
- Variation of Brownian paths, 262–263
- Variation of sample paths for processes with independent increments, 315–316
- Versions of stochastic processes, 257
- Ville, J., 103
- Volkonskiĭ, V., 390
- Wald's identity, 100
  - for Brownian motion, 275
- Wax, N., 271
- Weak convergence of measures, compactness property, 217
  - defined, 217
  - conditions for, 218
- Weyl, H., 117
- Widder, D., 184
- Wiener, N., 246, 247, 271
- Whitman, W., 120
- Wintner, A., 49
- Yaglom, A., 247
- Zeros of Brownian motion, 267–268
- Zygmund, A., 271